Inducing Approximately Optimal Flow Using Truthful Mediators

Ryan Rogers Aaron Roth* Jonathan Ullman† Zhiwei Steven Wu

May 11, 2015

Abstract

We revisit a classic coordination problem from the perspective of mechanism design: how can we coordinate a social welfare maximizing flow in a network congestion game with selfish players? The classical approach, which computes tolls as a function of known demands, fails when the demands are unknown to the mechanism designer, and naively eliciting them does not necessarily yield a truthful mechanism. Instead, we introduce a *weak mediator* that can provide suggested routes to players and set tolls as a function of reported demands. However, players can choose to ignore or misreport their type to this mediator. Using techniques from differential privacy, we show how to design a weak mediator such that it is an asymptotic ex-post Nash equilibrium for all players to truthfully report their types to the mediator and faithfully follow its suggestion, and that when they do, they end up playing a nearly optimal flow. Notably, our solution works in settings of incomplete information even in the absence of a prior distribution on player types. Along the way, we develop new techniques for privately solving convex programs which may be of independent interest.

*University of Pennsylvania Department of Computer and Information Sciences. Partially supported by an NSF CAREER award, NSF Grants CCF-1101389 and CNS-1065060, and a Google Focused Research Award. aaroth@cis.upenn.edu

†Columbia University Department of Computer Science. Supported by a Junior Fellowship from the Simons Society of Fellows. Email: jullman@cs.columbia.edu.

Contents

1 Introduction

Large, atomic traffic routing games model the common scenario in which n agents (say, residents of a city) must choose paths in some graph (the road network) to route a unit of flow (drive to work) between their target source/sink pairs. In aggregate, the decisions of each of these agents cause congestion on the edges (traffic), and each agent experiences a cost equal to the sum of the latencies of the edges she traverses, given the decisions of everyone else. The latencies on each edge are a function of the congestion on that edge.

This widely studied class of games presents several well known challenges:

1. First, for the social welfare objective, the price of anarchy is unboundedly large when the latencies can be arbitrary convex functions.
2. Second, in atomic routing games, equilibria are not unique, and hence equilibrium selection is an important problem.
3. Finally, as in most large games, players will be generally unaware of the types of their opponents, and so it is important to understand these games in settings of incomplete information.

One way to address the first challenge is to introduce carefully selected *tolls* on the edges, which modifies the game and decreases the price of anarchy. Indeed, so called *marginal cost tolls* make the socially optimal routing a Nash equilibrium. The marginal cost toll on each edge charges each agent the cost that she imposes on all other agents. However, in atomic congestion games with marginal cost tolls, the socially optimal routing is not necessarily the only Nash equilibrium routing, and so the price of anarchy can be larger than 1, and the coordination problem is still not solved. Moreover, because it is difficult to charge agents tolls *as a function of what others are doing* (as the marginal cost tolls do), there is a large literature that considers the problem of finding *fixed* tolls that induce the optimal routing, under various conditions Cole et al. (2003); Fleischer et al. (2004); Karakostas and Kolliopoulos (2004); Fleischer (2005); Swamy (2007); Fotakis et al. (2010)

This literature, however, assumes the agents' source/sink pairs are known, and computes the tolls as a function of this information. In this paper we instead take a mechanism design approach—the demands of the agents must be elicited, and agents may misrepresent their demands if it is advantageous to do so. Compared to standard mechanism design settings, our mechanism is somewhat restricted: it can only set anonymous tolls, and cannot require direct payments from the agents, and it also cannot force the agents to take any particular route. Because of these limitations, standard tools like the VCG mechanism do not apply. Instead, we approach the problem by introducing a *weak mediator* which also solves the 2nd and 3rd problems identified above—it solves the equilibrium selection problem, even in settings of incomplete information. The solutions we give are all *approximate* (both in terms of the incentives we guarantee, and our approximation to the optimal social welfare), but the solution approaches perfect as the game grows large.

Informally, a weak mediator is an intermediary with whom agents can choose to interact with. This leads to a new *mediated game*, related to the original routing game. In our setting, the weak mediator elicits the types of each agent. Based on the agents reports, it fixes constant tolls to charge on each edge, and then suggests a route for each agent to play. However, agents are free to act independently of the mediator. They need not report their type to it honestly, or even report a type at all. They are also not obligated to follow the route suggested by the mediator, and can deviate from it in arbitrary ways. Our goal is to design a mediator that incentivizes "good behavior" in the mediated game—that agents should truthfully report their type to the mediator, and then faithfully follow its suggestion. Moreover, we want that when agents do this, the resulting routing will be socially optimal.

Our main result is that this is possible in large routing games with convex loss functions. By large, we mean both that the number of players n is large, and that the latency functions are Lipschitz continuous—i.e. that no single agent can substantially affect the latency of any edge via a unilateral deviation. We give a weak mediator that makes "good behavior" an approximate *ex-post* Nash equilibrium—i.e. a Nash equilibrium in *every* game that might be induced by realizations of the agents types. This is an extremely robust solution concept that applies even when agents have no distributional knowledge of each other's types. In the limit as n goes to infinity, the approximate equilibrium becomes exact. The mediator also implements an approximately optimal routing, in that the welfare of the suggested routing is suboptimal by an additive term that is sublinear in n. Hence, if the cost of the optimal routing grows linearly, or nearly linearly in n, then the approximately optimal flow achieves a fraction of the optimal social welfare that is arbitrarily close to 1.

1.1 Our Techniques and Main Results

At a high level, the approach we take is to design a mediator which takes as input the reported source/destination pairs of each agent, and as a function of those reports:

1. Computes the optimal routing given the reported demands, and
2. Computes fixed tolls that make this routing a Nash equilibrium, and finally
3. Suggests to each player that they play their part of this optimal routing.

However, implementing each of these steps straightforwardly does not make good behavior an equilibrium in general. Agents may hope to gain in two ways by misreporting their type: they may hope to change the tolls charged on the path that they eventually take, and they may hope to change the algorithm's suggestions to other players, to change the edge congestions. Simply because the game is large, and hence each player has little direct effect on the costs of other players does not necessarily mean that no player's report can have large effect on an *algorithm* which is computing an equilibrium (see e.g. Kearns et al. (2014) for an example).

To address this problem, we follow the approach taken in Kearns et al. (2014); Rogers and Roth (2014) and compute the optimal routing and tolls using *joint differential privacy*. Informally, joint differential privacy guarantees that if any agent unilaterally misreports her demand, then it has only a small effect on the routes taken by *every other agent*, as well as on the tolls. (It of course has a very large effect on the route suggested to that agent herself, since she is always given a route between her reported source/sink pairs!) As we show, this is sufficient to guarantee that an agent cannot benefit substantially by misreporting her demand. Assuming the other agents behave honestly—meaning they report their true demand and follow their suggested route—then the fact that the algorithm *also* is guaranteed to compute a routing which forms an approximate equilibrium of the game, given the tolls, guarantees that agents cannot do substantially better than also playing honestly, and playing their part of the computed equilibrium.

In order to do this, we need to develop new techniques for convex optimization under joint differential privacy. In particular, in order to find the socially optimal flow privately, we need the ability to privately solve a convex program with an objective that is not linearly separable among players, and hence one for which existing techniques Hsu et al. (2014b) do not apply.

We now informally state the main theorem of this paper. It asserts that there is a mediator that incentivizes good behavior as an ex-post Nash equilibrium, while implementing the optimal flow. Here we assume that the latency functions on the edges are bounded by the number of players n

and are Lipschitz continuousalthough our formal theorem statement gives more general parameter tradeoffs.

Theorem 1.1 (Informal)**.** *For large*[1] *routing games with n players and m edges, there exists a mediator M such that good behavior is an η_{eq}-approximate Nash equilibrium in the mediated game where*

$$\eta_{eq} = \tilde{O}\left(m^{3/2} n^{4/5}\right)$$

and when players follow good behavior, the resulting flow is an η_{opt}-approximately optimal average flow for the original routing game where

$$\eta_{opt} = \tilde{O}\left(mn^{4/5}\right).$$

To interpret this theorem, let us write OPT to denote the *average player latency* in the socially optimal flow. Note that in this parameter regime (latency functions which are bounded by n and Lipschitz), if the value OPT increases at a rate faster than $n^{4/5}$ as the population n grows, then our mediator yields a flow that obtains average latency $(1 + o_n(1)) \cdot \text{OPT}$.[2] We view this condition on OPT as very mild. For example, if the network is fixed and all of the latency functions have derivatives bounded strictly away from zero, then the optimal average latency will grow at a rate of $\Omega(n)$. Our results hold *even* when the optimal average latency grows sublinearly. Similarly, in this setting, for a $1 - o_n(1)$ fraction of individuals the latency of their best response route also grows at a rate of $\Omega(n)$, and hence our mediator guarantees that for a $(1 - o_n(1))$-fraction of individuals, they are playing an $(1 - o_n(1))$-approximate best-response (i.e. they cannot decrease their latency by more than a $1 - o_n(1)$ multiplicative factor by deviating from the mediator's suggestion).

1.2 Related Work

There is a long history of using tolls to modify the equilibria in congestion games (see e.g. Beckmann et al. (1956) for a classical treatment). More recently, there has been interest in the problem of computing fixed tolls to induce optimal flows at equilibrium in various settings, usually in *non-atomic* congestion games (see e.g. Cole et al. (2003); Fleischer et al. (2004); Karakostas and Kolliopoulos (2004); Fleischer (2005); Swamy (2007); Fotakis et al. (2010) for a representative but not exhaustive sample). These papers study variations on the problem in which e.g. tolls represent lost welfare Cole et al. (2003), or in which agents have heterogenous values for money Fleischer et al. (2004), or when agents are atomic but flow is splittable Swamy (2007), among others. Tolls in atomic congestion games have received some attention as well (e.g. Caragiannis et al. (2006)), though to a lesser degree, since in general atomic congestion games, tolls do not suffice to implement the optimal flow as the unique equilibrium). These works all assume that agent demands are known, and do not have to be elicited from strategic agents, which is where the present paper departs from this literature. Recently, Bhaskar et al. Bhaskar et al. (2014) consider the problem of computing tolls in a query model in which the latency functions are unknown (demands are known), but not in a setting in which agents are assumed to be behaving strategically to manipulate the tolls.

Modifying games by adding "mediators" is also well studied, although what exactly is meant by a mediator differs from paper to paper (see e.g. Monderer and Tennenholtz (2003, 2009); Rozenfeld and Tennenholtz (2007); Ashlagi et al. (2009); Peleg and Procaccia (2010) for a representative but not exhaustive sample). The "weak mediators" we study in this paper were introduced in

[1] The formal notion of largeness we require is detailed in Assumption 2.3.

[2] Here, $o_n(1)$ denotes a function of n that approaches 0 as $n \to \infty$.

Kearns et al. (2014); Rogers and Roth (2014), who also use differentially private equilibrium computation to achieve incentive properties. Our work differs from this prior work in that Kearns et al. (2014); Rogers and Roth (2014) both seek to implement an equilibrium of the given game, and hence do not achieve welfare guarantees beyond the price of anarchy of the game. In contrast, we use tolls to modify the original game, and hence implement the socially optimal routing as an equilibrium.

The connection between differential privacy, defined by Dwork et al. (2006), and mechanism design was first made by McSherry and Talwar (2007), who used it to give improved welfare guarantees for digital goods auctions. It has since been used in various contexts, including to design mechanisms for facility location games and general mechanism design problems without money Nissim et al. (2012). The connection between *joint* differential privacy and mechanism design (which is more subtle, and requires that the private algorithm also compute an equilibrium of some sort) was made by Kearns et al. (2014) in the context of mediators, and has since been used in other settings including computing stable matchings Kannan et al. (2015), aggregative games Cummings et al. (2014), and combinatorial auctions Hsu et al. (2014b).

2 Model

2.1 The Routing Game Problem

In this section we introduce the atomic unsplittable routing game problem that we study. An instance of a routing game $\Gamma = (G, \ell, \mathbf{s})$ is defined by

- A graph $G = (V, E)$. We use $m = |E|$ to denote the number of edges.
- A *latency function* $\ell_e : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ for each edge $e \in E$. Each latency function maps the number of players who send flow along that edge to a non-negative loss.
- A set of n source-destination pairs $\mathbf{s} = (s_1, \ldots, s_n)$. Each pair $s_i = (s_i^1, s_i^2) \in \mathcal{S} \equiv V \times V$ represents the *demand of player* i. We use n to denote the number of players.

The objective is to (approximately) minimize the total latency experienced by all the players in the network. Let $\mathcal{F}(\mathbf{s}) = (\mathcal{F}(s_1), \cdots, \mathcal{F}(s_n))$ be the set of *feasible individual flows for demand* $\mathbf{s}$ and $\mathcal{F} = \{\mathcal{F}(\mathbf{s}) : \mathbf{s} \in \mathcal{S}\}$ be the set of all feasible individual flows. Notice that an element of $\mathcal{F}(\mathbf{s})$ is a vector of n separate flows, one for each player. That is, an individual flow is specified by $n \times m$ variables representing the amount of flow by each player routed on each edge. Specifically, given a graph G, $\mathcal{F}(\mathbf{s})$ is the set of unsplittable flows $\mathbf{x} = (x_{i,e})_{i \in [n], e \in E} \in \{0,1\}^{n \times m}$ such that

$$b_{i,u} = \begin{cases} 1 & u = s_i^1 \\ -1 & u = s_i^2 \\ 0 & \text{else} \end{cases} \tag{1}$$

$$b_{i,u} + \sum_{v:(u,v)\in E} x_{i,(u,v)} = \sum_{v:(v,u)\in E} x_{i,(v,u)}, \qquad \forall u \in V \quad \forall i \in [n] \tag{2}$$

For a given routing game instance $\Gamma = (G, \ell, \mathbf{s})$, we seek a flow $\mathbf{x} \in \mathcal{F}(\mathbf{s})$ that minimizes the average latency $\phi(\mathbf{x})$

$$\phi(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^{n} \sum_{e \in E} x_{i,e} \cdot \ell_e \left(\sum_{i=1}^{n} x_{i,e} \right) \tag{3}$$

We will sometimes write $\mathrm{OPT}(\mathbf{s}) = \phi(\mathbf{x}^*)$, were $\mathbf{x}^*$ is the minimum average cost flow for the routing game $\Gamma = (G, \ell, \mathbf{s})$ when the graph G and latencies ℓ are known from context. In this work we settle for an approximately minimum average cost flow, which we define below.

Definition 2.1 (Approximately Optimal Flow)**.** For a routing game Γ, and parameter $\eta_{opt} > 0$, a flow $\mathbf{x}$ is *η_{opt}-approximately optimal* if $\mathbf{x} \in \mathcal{F}(\mathbf{s})$ and

$$\phi(\mathbf{x}) \leq \mathrm{OPT}(\mathbf{s}) + \eta_{opt}.$$

We are interested in strategic players that want to minimize their individual cost

$$\phi_i(\mathbf{x}) = \sum_{e \in E} x_{i,e} \cdot \ell_e \left(\sum_{j=1}^{n} x_{j,e} \right). \tag{4}$$

We thus define an approximate Nash flow.

Definition 2.2 (Approximate Nash Flow)**.** For a routing game Γ and parameter $\eta_{eq} > 0$, a flow $\hat{\mathbf{x}}$ is an *η_{eq}-approximate Nash flow* if $\hat{\mathbf{x}} \in \mathcal{F}(\mathbf{s})$ and for every $x_i \in \mathcal{F}(s_i)$

$$\phi_i(\hat{\mathbf{x}}) \leq \phi_i(x_i, \hat{\mathbf{x}}_{-i}) + \eta_{eq} \qquad \forall x_i \in \mathcal{F}(s_i).$$

When $\hat{\mathbf{x}}$ is a 0-approximate Nash flow, we simply say that it is a *Nash flow.*

Throughout, we will make the following assumptions about the latency functions.

Assumption 2.3. For every edge $e \in E$, the latency function ℓ_e is (1) non-decreasing, (2) convex, (3) twice differentiable, (4) bounded by n (i.e. $\ell_e(n) \leq n$), and (5) γ-lipschitz (i.e. $|\ell_e(y) - \ell_e(y')| \leq \gamma |y - y'|$ for all $e \in E$) for some constant $\gamma > 0$.

Item 1 and 2 are natural and extremely common in the routing games literature. Item 3 is a technical condition used in our proofs that can likely be removed. Item 4 and 5 are the "largeness conditions" that ensure no player has large influence on any other's payoff. If the Lipschitz constant is zero, then we can choose an upper bound parameter $\gamma > 0$ in our analysis.

2.2 Mediators

Given an instance $\Gamma = (G, \ell, \mathbf{s})$, we would like the players to coordinate on the social-welfare maximizing flow $\mathbf{x}^*$ where $\mathrm{OPT}(\mathbf{s}) = \phi(\mathbf{x}^*)$. There are two problems: the first is that the optimal flow is generally not a Nash equilibrium, and the second is that even with knowledge of everyone's demands, Nash equilibria are not unique and coordination is a problem. The classical solution to the first problem is to have an overseer impose edge tolls τ, which are a function of the demands $\mathbf{s}$ of each player. This makes $\mathbf{x}^*$ a Nash flow for the routing game instance $\Gamma^\tau = (G, \ell^\tau, \mathbf{s})$ where

$$\ell_e^\tau(y) = \ell_e(y) + \tau_e.$$

However the tolls that cause the optimal flow to be an equilibrium depend on the demands, and so this approach fails if the overseer does not know $\mathbf{s}$. A simple solution would be to elicit the demands from the players, but since the correct tolls depend on the demands, naively eliciting them may not lead to a truthful mechanism.

We solve this problem, as well as the equilibrium selection problem mentioned above, by introducing a *mediator* that takes as input the demand of each player and outputs a set of tolls for each edge, together with a suggested route for each player to use. Ideally, the players will report their demands truthfully, the aggregate of the routes suggested by the mediator will be a social-welfare maximizing flow $\mathbf{x}^*$, agents will faithfully follow their suggestion, and the tolls will be chosen to make $\mathbf{x}^*$ an (approximate) Nash flow. However, players have the option to deviate from this desired behavior in several ways: they may not report their demand to the mediator at all, might report a false demand, or might not follow the mediator's suggestion once it is given. Our goal in designing the mediator is to guarantee that players never have significant incentive to deviate from the desired behavior described above.

Formally, introducing the mediator gives rise to a modified game $\Gamma_M = (G, \ell, \mathbf{s}, M)$. The mediator is an algorithm $M : \{\perp \cup \mathcal{S}\}^n \to \mathcal{F}^n \times \mathbb{R}^m$. The input from each player is either a demand or a $\perp$ symbol indicating that the player opts out. The output is a set of routes, one suggested to each player, together with a collection of tolls, one for each edge. We write the output as

$$M(\mathbf{s}) = \left(\left(M_i^{\mathcal{F}}(\mathbf{s})\right)_{i \in [n]}, M^{\tau}(\mathbf{s}) \right).$$

The edge tolls $M^{\tau}(\mathbf{s}) = (M_e^{\tau}(\mathbf{s}))_{e \in E}$ that M outputs will enforce the optimal flow induced by the reported demands. Note that the tolls that M outputs to each player are the same (i.e. the players are not charged personalized tolls; rather there is a single toll on each edge that must be paid by any player using that edge).

In Γ_M each player can *opt-out* of using the mediator, denoted by the report $\perp$, and then select some way to route from his source to his destination, or a player can *opt-in* to using the mediator, but not necessarily reveal her true demand, and then the mediator will suggest a path $\mathbf{x}_i$ to route her unit flow from the reported source to the destination. Players are free to follow the suggested action, but they can also use the suggestion as part of an arbitrary deviation, i.e. they can play any action $f(\mathbf{x}_i)$ for any $f : \mathcal{F} \to \mathcal{F}$. Thus, the action set $\mathcal{A}$ for any player for the game instance Γ_M is $\mathcal{A} = A_1 \cup A_2$ where $A_1 = \{(s', f) : s' \in \mathcal{S}, f : \mathcal{F} \to \mathcal{F}\}$ and $A_2 = \{(\perp, f) : f \text{ constant }\}$.

We next define the cost function for each player in Γ_M, but first we must present some notation. Let $\mathbf{F}$ be the set of possible functions $f_i : \mathcal{F} \to \mathcal{F}$, where $f_i(\mathbf{x}_i) = (f_{i,e}(x_{i,e}))_{e \in E}$. We further write $f_e(\mathbf{x}) = \sum_{i=1}^n f_{i,e}(x_{i,e})$ as the new congestion on edge e when players have deviated from $\mathbf{x}$ according to functions f_i for $i \in [n]$. We will consider only randomized algorithms, so our cost is an expectation over outcomes of M. More formally, the cost ϕ^M that each player experiences in Γ_M is defined as

$$\begin{aligned} &\phi^M : \mathcal{S} \times [(\perp \cup \mathcal{S}) \times \mathbf{F}]^n \to \mathbb{R} \\ &\phi^M(s_i, (\mathbf{s}', \mathbf{f})) := \mathop{\mathbb{E}}_{(\mathbf{x},\tau) \sim M(\mathbf{s}')} \left[\sum_{e \in E} f_{i,e}(x_{i,e}) \left(\underbrace{\ell_e\left(f_e(\mathbf{x})\right) + \tau_e}_{\ell_e^{\tau}(f_e(\mathbf{x}))} \right) \right] \end{aligned}$$

where s_i is player i's true source-destination pair.

We are interested in designing mediators such that *good behavior* in the mediated game is an ex-post Nash equilibrium, which we define below.

Definition 2.4 (Ex-Post Nash Equilibrium)**.** A set of strategies $\{\sigma_i \colon \mathcal{S} \to \mathcal{A}\}_{i=1}^n$ forms an η-approximate ex-post Nash equilibrium if for every profile of demands $\mathbf{s} \in \mathcal{S}^n$, and for every player i and action $a_i \in \mathcal{A}$:

$$\phi^M\left(s_i, (\sigma_i(s_i), \sigma_{-i}(s_{-i}))\right) \le \phi^M\left(s_i, (a_i, \sigma_{-i}(s_{-i}))\right) + \eta.$$

That is, it forms an η-approximate Nash equilibrium for every realization of demands.

Our goal is to incentivize players to follow *good behavior*—truthfully reporting their demand, and then faithfully following the suggested action of the mediator. Formally, the good behavior strategy for player i is $\xi_i(s_i) = (s_i, \mathrm{id})$ where s_i is i's actual demand, and $\mathrm{id} \colon \mathcal{F} \to \mathcal{F}$ is the identity map. We write $\xi_i = \xi_i(s_i)$ for the good behavior strategy.

To accomplish this goal, we will design a mediator that is "insensitive" to the reported demand of each player. Informally, if a player's reported demand does not substantially effect the tolls chosen by the mediator, or the paths suggested to *other* players, then a player has little incentive to lie about his demand (of course any mediator with this property must necessarily allow the path suggested to agent i to depend strongly on agent i's own reported demand!). We capture this notion of insensitivity using *joint differential privacy* Kearns et al. (2014), which is defined as follows.

Definition 2.5. (Joint Differential Privacy Kearns et al. (2014)) A randomized algorithm $\mathcal{M} : \mathcal{S}^n \to \mathcal{O}^n$, where $\mathcal{O}$ is an arbitrary output set for each player, satisfies (ε, δ)-joint differential privacy if for every player i, every pair $s_i, s_i' \in \mathcal{S}$, any tuple $s_{-i} \in \mathcal{S}^{n-1}$ and any $B_{-i} \subseteq \mathcal{O}^{n-1}$, we have $\mathbb{P}\left[\mathcal{M}(s_i, s_{-i})_{-i} \in B_{-i}\right] \le e^{\varepsilon} \cdot \mathbb{P}\left[\mathcal{M}(s_i', s_{-i})_{-i} \in B_{-i}\right] + \delta$.

Joint differential privacy (JDP) is a relaxation of the notion of *differential privacy* (DP) Dwork et al. (2006). We state the definition of DP below, both for comparison, and because it will be important technically in designing our mediator.

Definition 2.6. (Differential Privacy Dwork et al. (2006)) A randomized algorithm $\mathcal{M} : \mathcal{S}^n \to \mathcal{O}$ satisfies (ε, δ)-differential privacy if for any player i, any two $s_i, s_i' \in \mathcal{S}$, any tuple $s_{-i} \in \mathcal{S}^{n-1}$, and any $B \subseteq \mathcal{O}$ we have $\mathbb{P}\left[\mathcal{M}(s_i, s_{-i}) \in B\right] \le e^{\varepsilon} \cdot \mathbb{P}\left[\mathcal{M}(s_i', s_{-i}) \in B\right] + \delta$.

Note that JDP is weaker than DP, because JDP assumes that the output space of the algorithm is partitioned among the n players, and the output to player i can depend arbitrarily on the input of player i, and only the output to players $j \neq i$ must be insensitive to the input of player i. This distinction is crucial in mechanism design settings—the output to player i is a suggested route for player i to follow, and thus should satisfy player i's reported demand, which is highly sensitive to the input of player i. Also note that since our mediator will output the same tolls to every player, the tolls computed by the mediator must satisfy standard DP.

A key property we use is that a JDP mediator that also computes an equilibrium of the underlying game gives rise to an approximately truthful mechanism. This result was first shown in Kearns et al. (2014); Rogers and Roth (2014), although for simpler models that do not include tolls. We now state and prove a simple extension of this result that is appropriate for our setting.

Theorem 2.7. *Given routing game $\Gamma = (G, \ell, \mathbf{s})$ and upper bound U on the tolls, let $M : (\bot \cup \mathcal{S})^n \to \mathcal{F}^n \times [0, U]^m$ where $M(\mathbf{s}') = \left(M_i^{\mathcal{F}}(\mathbf{s}'), M^{\tau}(\mathbf{s}')\right)_{i \in [n]}$ satisfies*

1. *M is (ε, δ)-joint differentially private.*
2. *For any input demand profile $\mathbf{s}$, we have with probability $1 - \beta$ that $\mathbf{x} = \left(M_i^{\mathcal{F}}(\mathbf{s})\right)_{i=1}^n$ is an η_{eq}-approximate Nash flow in the modified routing game $\Gamma^{\tau} = (G, \ell^M, \mathbf{s})$ where*

$$\ell_e^M(y) := \ell_e(y) + M_e^{\tau}(\mathbf{s}) \quad \forall e \in E.$$

Then the good behavior strategy $\xi = (\xi_1, \ldots, \xi_n)$ forms an η-approximate ex-post Nash equilibrium in $\Gamma_M = (G, \ell, \mathbf{s}, M)$, where

$$\eta = \eta_{eq} + m(U + n)(2\varepsilon + \beta + \delta).$$

Proof. We fix $\mathbf{s} \in \mathcal{S}^n$ to be the true source destination of the players. We consider a unilateral deviation $\xi_i'(s_i) = (s_i', f_i')$ for player i to report s_i' and use f_i', which we write as ξ_i'. We write the modified cost function for player i in Γ^τ with tolls $\tau_e = M_e^\tau(\mathbf{s})$ to be

$$\phi_i^\tau(x_i, \mathbf{x}_{-i}) = \sum_{e \in E} x_{i,e} \left(\ell_e \left(\sum_{j=1}^{n} x_{j,e} \right) + \tau_e \right)$$

We define the best response flow that player i of demand s_i can route given the flows of the other players to be

$$BR_i^\tau(\mathbf{x}_{-i}) = \underset{x_i \in \mathcal{F}(s_i)}{\text{argmin}} \left\{ \phi_i^\tau(x_i, \mathbf{x}_{-i}) \right\}.$$

We first condition on the event that M gives an η_{eq}-approximate Nash flow in Γ^τ.

$$\phi^M(s_i, (\xi_i, \xi_{-i})) = \underset{(\mathbf{x},\tau)\sim M(\mathbf{s})}{\mathbb{E}} \left[\phi_i^\tau(x_i, \mathbf{x}_{-i}) \right] \leq \underset{(\mathbf{x},\tau)\sim M(\mathbf{s})}{\mathbb{E}} \left[\phi_i^\tau(BR_i^\tau(\mathbf{x}_{-i}), \mathbf{x}_{-i}) \right] + \eta_{eq}$$

We then use the fact that M is JDP. We write $\mathbf{s}' = (s_i', \mathbf{s}_{-i})$.

$$\begin{aligned}
\phi^M(s_i, (\xi_i, \xi_{-i})) &\leq e^\varepsilon \left(\underset{(\mathbf{x},\tau)\sim M(\mathbf{s}')}{\mathbb{E}} \left[\phi_i^\tau(BR_i^\tau(\mathbf{x}_{-i}), \mathbf{x}_{-i}) \right] \right) + m(U+n)\delta + \eta_{eq} \\
&\leq \underset{(\mathbf{x},\tau)\sim M(\mathbf{s}')}{\mathbb{E}} \left[\phi_i^\tau(BR_i^\tau(\mathbf{x}_{-i}), \mathbf{x}_{-i}) \right] + m(U+n)\left(2\varepsilon + \delta\right) + \eta_{eq} \\
&\leq \underset{(\mathbf{x},\tau)\sim M(\mathbf{s}')}{\mathbb{E}} \left[\phi_i^\tau(f_i'(x_i), \mathbf{x}_{-i}) \right] + m(U+n)\left(2\varepsilon + \delta\right) + \eta_{eq}
\end{aligned}$$

The first inequality comes from using the fact that M is (ε, δ)-JDP and the fact that $\phi_i^\tau(\mathbf{x}) \leq m(U+n)$. The second inequality uses the fact that $e^\varepsilon \leq 1 + 2\varepsilon$ for $\varepsilon < 1$. The last inequality follows from the fact that player i can only do worse by not best responding to the other players' flows. Lastly, we know that M does not produce an η_{eq}-approximate Nash flow in Γ^τ with probability less than β, which gives the additional β term in the theorem statement. □

The rest of the paper will be dedicated to constructing such a mediator that satisfies the hypotheses in Theorem 2.7. We now state the main result of our paper.

Theorem 2.8. *For routing games Γ that satisfy Assumption 2.3 and parameter $\beta > 0$, there exists a mediator $M : \{\perp \cup \mathcal{S}\}^n \to \mathcal{F}^n \times [0, n\gamma]^m$ such that with probability $1 - \beta$ good behavior forms an η-approximate ex-post Nash equilibrium in Γ_M where*

$$\eta = \tilde{O}\left(m^{3/2} n^{4/5}\right)$$

and the resulting flow from the good behavior strategy is η_{opt}-approximately optimal for

$$\eta_{opt} = \tilde{O}\left(m n^{4/5}\right).$$

3 Flow Mediator with Tolls

We start by presenting a high level overview of the design of our algorithm. Our goal is to design a mediator that takes as input the demands, or source-destination pairs, $\mathbf{s}$ of the players and outputs a nearly optimal flow $\mathbf{x}^\bullet$ for $\Gamma = (G, \ell, \mathbf{s})$ together with edge tolls τ, such that the tolls are not heavily influenced by any single player's report and no one's report has major influence on the flow induced by the other players. Further, we need the tolls τ to be carefully computed so that $\mathbf{x}^\bullet$ is also an approximate Nash flow in the instance $\Gamma^\tau = (G, \ell^\tau, \mathbf{s})$. We construct such a mediator in the following way:

1. We compute an approximately optimal flow $\mathbf{x}^\bullet$ subject to JDP, using a privacy preserving variant of projected gradient descent. This ends up being the most technical part of the paper and so we leave the details to Section 4 and give the formal algorithm `P-GD` in Algorithm 6.For the rest of this section we assume we have $\mathbf{x}^\bullet$.

2. Given $\mathbf{x}^\bullet$, we need to compute the necessary tolls $\hat{\tau}$ such that players are approximately best responding in $\Gamma^{\hat{\tau}} = (G, \ell^{\hat{\tau}}, \mathbf{s})$ when playing $\mathbf{x}^\bullet$. We compute $\hat{\tau}$ as a function of a noisy version of the edge congestion $\hat{\mathbf{y}}$ induced by the flow $\mathbf{x}^\bullet$ so that $\hat{\tau}$ is DP. We give the procedure `P-CON` that computes $\hat{\mathbf{y}}$ in Algorithm 2. We must be cautious at this step because $\mathbf{x}^\bullet$ is only approximately optimal (and the tolls are computed with respect to a perturbed version of the induced congestion), so there may be a few players that are not playing approximate best responses in $\Gamma^{\hat{\tau}}$. We call these players *unsatisfied*.

3. We show that the number of *unsatisfied* players in $\Gamma^{\hat{\tau}}$ with flow $\mathbf{x}^\bullet$ is small, so we can modify $\mathbf{x}^\bullet$ by having the unsatisfied players play best responses to the induced flow. Because the number of unsatisfied players was small, we can show that this modification does not substantially reduce the payoff of the other players. Therefore, if those players were playing approximate best responses before the modification, they will continue to do so after. The procedure `P-BR`, given in Algorithm 3, ensures every player is approximately best responding. The result is a slightly modified flow $\hat{\mathbf{x}}$ which is nearly optimal in Γ and an approximate Nash flow in $\Gamma^{\hat{\tau}}$.

4. The final output is then $\hat{\mathbf{x}}$ and $\hat{\tau}$.

Our mediator `FlowToll` is formally given in Algorithm 1 and is composed of the subroutines described above. In `FlowToll` we are using `P-GD` as a black box that computes an α-approximate optimal flow. Theorem 4.10 shows that we can set

$$\alpha = \tilde{O}\left(\frac{\sqrt{n}m^{5/4}}{\sqrt{\varepsilon}}\right). \tag{5}$$

The rest of this paper is dedicated to analyzing the subroutines of `FlowToll`.

Remark 3.1. *Throughout our discussion of the subroutines, we will sometimes say "player i plays..." or "player i best responds to..." to describe player i's action in some flow computed by these subroutines. While these descriptions are natural, they could be slightly misleading. We want to clarify that our mediator mechanism is not interactive or online, and all the computation is done by the algorithm. The players will simply submit their private source-destination pairs and will only receive a suggested feasible path along with the tolls over the edges.*

```
Algorithm 1 Flow Mediator with Tolls
Input: A routing game instance $\Gamma = (G, \ell, \mathbf{s})$; privacy parameter $(\varepsilon, \delta)$; failure probability $\beta$
Output: $\hat{x}_i$, a $(s_i^1, s_i^2)$-flow for each player $i \in [n]$, and a toll $\hat{\tau}_e$ for each edge $e \in E$
  procedure FlowToll($\Gamma, \varepsilon, \delta, \beta$)
      1. Compute an $\alpha$-approximately optimal flow
             $\mathbf{x}^\bullet \leftarrow \texttt{P-GD}\left(\Gamma, \frac{\varepsilon}{4}, \frac{\delta}{2}, \frac{\beta}{2}\right).$
      2. Compute congestion $\hat{\mathbf{y}} \leftarrow \texttt{P-CON}(\mathbf{x}^\bullet, \varepsilon/4)$ and tolls $\hat{\tau} \leftarrow \tau^*(\hat{y}_e)$ where $\tau^*(\cdot)$ is given in (6).
      3. Improve some players' paths
             $\hat{\mathbf{x}} \leftarrow \texttt{P-BR}\left(\Gamma^{\hat{\tau}}, \hat{\mathbf{y}}, \mathbf{x}^\bullet, 4\sqrt{m\gamma\alpha} + \frac{8\gamma m^2 \log(2m/\beta)}{\varepsilon}\right).$
  return $\hat{\mathbf{x}}$ and $\hat{\tau}$
  end procedure
```

3.1 Private Tolls Mechanism

We show in this section that given an approximately optimal flow $\mathbf{x}^\bullet$ we can compute the necessary tolls $\hat{\tau}$ in a DP way. Ultimately, we want to compute *constant* tolls, but a useful intermediate step is to consider the following *functional* tolls, which are edge tolls that can depend on the congestion on that edge. Specifically, we define the *marginal-cost toll* $\tau_e^* \colon \mathbb{R} \to \mathbb{R}$ for each edge $e \in E$ to be

$$\tau_e^*(y) = (y-1)(\ell_e(y) - \ell_e(y-1)), \tag{6}$$

which gives rise to a different routing game $\Gamma^{\tau^*} = (G, \ell^{\tau^*}, \mathbf{s})$ with latency function defined as $\ell_e^{\tau^*}(y) = \ell_e(y) + \tau_e^*(y)$ for $e \in E$.

We first show that a marginal-cost toll enforces the optimal flow in an atomic, unsplittable routing game, and then show how to use this fact to privately compute *constant* tolls that approximately enforce the optimal flow at equilibrium. Recall the classical potential function method Monderer and Shapley (1996) for congestion games that defines a potential function $\Psi : \mathbb{R}^{n\times m} \to \mathbb{R}$ such that a flow $\mathbf{x}$ that minimizes Ψ is also a (exact) Nash flow in $\Gamma^{\tau^*} = (G, \ell^{\tau^*}, \mathbf{s})$, where

$$\Psi(\mathbf{x}) := \sum_{e\in E}\sum_{i=1}^{y_e} \ell_e^{\tau^*}(i) = \sum_{e\in E}\sum_{i=1}^{y_e} \left[\ell_e(i) + \tau_e^*(i)\right], \quad \text{and } y_e = \sum_{i\in[n]} x_{i,e}. \tag{7}$$

Lemma 3.2. *Let $\mathbf{x}^*$ be the (exact) optimal flow in routing game $\Gamma = (G, \ell, \mathbf{s})$, then $\mathbf{x}^*$ is a Nash flow in $\Gamma^{\tau^*} = (G, \ell^{\tau^*}, \mathbf{s})$*

Proof. First, we show that $n \cdot \phi(\mathbf{x}) = \Psi(\mathbf{x})$ where ϕ is given in (3):

$$\begin{aligned}\Psi(\mathbf{x}) &= \sum_{e}\sum_{i=1}^{y_e}\left[\ell_e(i) + \tau_e^*(i)\right] = \sum_{e}\sum_{i=1}^{y_e}\left[\ell_e(i) + (i-1)(\ell_e(i) - \ell_e(i-1))\right] \\ &= \sum_{e}\sum_{i=1}^{y_e}\left[i\,\ell_e(i) - (i-1)\,\ell_e(i-1)\right] = \sum_{e}\left[y_e\,\ell_e(y_e) - 0\,\ell_e(0)\right] = \sum_{e} y_e\,\ell_e(y_e) = n\cdot\phi(\mathbf{x}).\end{aligned}$$

Note that $\mathbf{x}^*$ minimizes the potential function Ψ. We know from Monderer and Shapley (1996) that the flow that minimizes the potential function Ψ is a Nash flow of the routing game Γ^{τ^*}. Hence $\mathbf{x}^*$ is a Nash flow. □

Since we only have access to an approximately optimal flow $\mathbf{x}^\bullet$, we will compute the marginal-cost tolls based on $\mathbf{x}^\bullet$ instead. In order to release DP tolls, we compute them using a private version $\hat{y}_e$ of the total edge congestion $y_e = \sum_i x^\bullet_{i,e}$ that is output by `P-CON` (presented in Algorithm 2). Using a standard technique in differential privacy, we can release a private version of the edge congestion by perturbing the congestion on each edge with noise from an appropriately scaled Laplace distribution. Since the analysis is standard, we defer the details to Section A.1. Lastly, to get the constant tolls for the mediator `FlowToll`, we will evaluate the marginal-cost toll function on the perturbed edge congestion $\hat{\mathbf{y}}$: set $\hat{\tau}_e = \tau^*_e(\hat{y}_e)$ for $e \in E$.

Algorithm 2 Private Congestion

```
Input: Flow $\mathbf{x}$, privacy parameter $\varepsilon$
Output: Aggregate flow $\hat{\mathbf{y}} = (\hat{y}_e)_{e \in E}$
  procedure P-CON($\mathbf{x}, \varepsilon$)
      for each edge $e \in E$ do
          Let $\hat{y}_e = \sum_i x_{i,e} + Z_e,$ where $Z_e \sim \text{Lap}(m/\varepsilon)$.
          if $\hat{y}_e > n$ then
              $\hat{y}_e \leftarrow n$.
      return $\hat{\mathbf{y}}$
  end procedure
```

To show that the constant tolls $\hat{\tau}$ are private, we need to first show that the noisy congestion $\hat{\mathbf{y}}$ output by `P-CON` is DP in the demands $\mathbf{s}$. We will show later that `P-GD` which computes $\mathbf{x}^\bullet$ is JDP in $\mathbf{s}$. We then use $\mathbf{x}^\bullet$ as input to `P-CON`, which we know is DP with respect to any flow input $\mathbf{x}$. To bridge the two privacy guarantees, we rely on the following composition lemma (with proof in Appendix A.3) to show that $\hat{\mathbf{y}}$ is DP in $\mathbf{s}$.

Lemma 3.3. *Let $M_J : \mathcal{S}^n \to \mathcal{X}^n$ be (ε_J, δ)-jointly differentially private. Further, let $M_D : \mathcal{X}^n \to O$ be ε_D-differentially private. If $M : \mathcal{S}^n \to O$ is defined as*

$$M(\mathbf{s}) = M_D(M_J(\mathbf{s}))$$

then M is $(2\varepsilon_D + \varepsilon_J, \delta)$-differentially private.

Now we are ready to establish the privacy guarantee of both $\hat{\mathbf{y}}$ and $\hat{\tau}$.

Corollary 3.4. *Given the approximately optimal flow $\mathbf{x}^\bullet$ computed from `P-GD`$(\Gamma, \varepsilon/4, \delta/2, \beta/2)$, the perturbed congestion $\hat{\mathbf{y}}$ output by `P-CON`$(\mathbf{x}^\bullet, \varepsilon/4)$ and the constant tolls $\hat{\tau} = (\tau^*_e(\hat{y}_e))_{e \in E}$ are $(3\varepsilon/4, \delta/2)$-differentially private in the demands $\mathbf{s}$.*

Proof. Note that $\mathbf{x}^\bullet$ is output by `P-GD`$(\Gamma, \varepsilon/4, \delta/2, \beta/2)$, so it is $(\varepsilon/4, \delta/2)$-JDP in $\mathbf{s}$. Using analysis of the Laplace mechanism(Section A.1), we know that `P-CON`$(\mathbf{x}^\bullet, \varepsilon/4)$ is $(\varepsilon/4)$-DP in $\mathbf{x}^\bullet$. Therefore, the noisy congestion $\hat{\mathbf{y}}$ output by the composition of these two functions is $(3\varepsilon/4, \delta/2)$-DP by Lemma 3.3. Since $\hat{\tau}$ is simply a post-processing of the noisy congestion $\hat{\mathbf{y}}$, we know that $\hat{\tau}$ is $(3\varepsilon/4, \delta/2)$-DPby Lemma A.1. □

3.2 Simultaneous Best Responses of Unsatisfied Players

At this point of the mechanism, we have computed the approximately optimal flow $\mathbf{x}^\bullet$ and constant tolls $\hat{\tau}$ that define the tolled routing game $\Gamma^{\hat{\tau}}$. In this section, we show how to modify $\mathbf{x}^\bullet$ to obtain a new approximately optimal flow $\hat{\mathbf{x}}$ that is also an approximate Nash equilibrium in the presence of the same constant tolls $\hat{\tau}$.

Recall from Lemma 3.2 that there is an exactly optimal flow $\mathbf{x}^*$ and functional tolls τ^* such that $\mathbf{x}^*$ is an exact Nash flow of the routing game under tolls τ^*. Our flow-toll pair $(\mathbf{x}^\bullet, \hat{\tau})$ differs from $(\mathbf{x}^*, \tau^*)$ in three ways.

1. The flow $\mathbf{x}^\bullet$ is only *approximately* optimal.

2. The tolls $\hat{\tau}$ we impose on the edges are *constants* while the functional tolls τ^* may be functions of the congestion.

3. Tolls $\hat{\tau}$ are derived from noisy congestion $\hat{\mathbf{y}}$, not the exact congestion $\mathbf{y}^\bullet = \sum_i \mathbf{x}_i^\bullet$.

As a result, there may be some *unsatisfied players* who could significantly benefit from deviating from $\mathbf{x}^\bullet$. We obtain the new approximate Nash flow $\hat{\mathbf{x}}$ by rerouting the unsatisfied players in $\mathbf{x}^\bullet$ along their best response route in the flow $\mathbf{x}^\bullet$ with constant edge tolls $\hat{\tau}$. To analyze the new flow $\hat{\mathbf{x}}$, we show that there are not too many unsatisfied players. Thus, even if we modify the routes of all of the unsatisfied players, the overall congestion does not change too much, and thus the players who were previously satisfied remain satisfied.

To determine if a player is unsatisfied and what their best response is, we need to know the costs they face for different paths, which depends on the flow $\mathbf{y}^\bullet = \sum_i \mathbf{x}_i^\bullet$. However, to ensure privacy, we only have access to a perturbed flow $\hat{\mathbf{y}}$. Thus, we will define unsatisfied players relative to this noisy flow $\hat{\mathbf{y}}$ computed by `P-CON`. More generally we can define the best response function of a player relative to any flow $\mathbf{y}$.

Given any congestion $\mathbf{y}$ (not necessarily even a sum of feasible individual flows) and routing game $\Gamma = (G, \ell, \mathbf{s})$, we define $c_{\mathbf{x}_i}(\mathbf{y})$ to be player i's cost for routing on path $\mathbf{x}_i$ under the congestion of $\mathbf{y}$, that is

$$c_{\mathbf{x}_i}(\mathbf{y}) = \sum_{e \in E} x_{i,e} \cdot \ell_e(y_e). \tag{8}$$

Note that $\sum_{i=1}^n c_{\mathbf{x}_i}(\mathbf{y}) = n\phi(\mathbf{x})$ and $c_{\mathbf{x}_i}(\mathbf{y}) = \phi_i(\mathbf{x})$ when $y_e = \sum_{i=1}^n x_{i,e}$ for $e \in E$. We then define the condition for being unsatisfied with respect to congestion $\mathbf{y}$ as follows.

Definition 3.5. Given congestion $\mathbf{y}$ and routing game $\Gamma = (G, \ell, \mathbf{s})$, we say that a player i with s_i-flow $\mathbf{x}_i$ is *ρ-unsatisfied with respect to* $\mathbf{y}$ if he could decrease his cost by at least ρ via a unilateral deviation. That is, there exists a path $\mathbf{x}'_i \in \mathcal{F}(s_i)$ such that

$$c_{\mathbf{x}'_i}(\mathbf{y}') \le c_{\mathbf{x}_i}(\mathbf{y}) - \rho$$

where $\mathbf{y}' = \mathbf{y} - \mathbf{x}_i + \mathbf{x}'_i$ is the flow that would result from player i making this deviation. If player i is not ρ-unsatisfied, then we say i is *ρ-satisfied.* We will sometimes omit $\mathbf{y}$ if it is clear from context.

The next lemma bounds the number of unsatisfied players in $\mathbf{x}^\bullet$ in the routing game $\Gamma^{\hat{\tau}} = (G, \ell + \hat{\tau}, \mathbf{s})$ with respect to the noisy congestion $\hat{\mathbf{y}}$.

Lemma 3.6. *Let $\mathbf{x}^\bullet$ be an α-approximately optimal flow, $\hat{\mathbf{y}} =$ `P-CON`$(\mathbf{x}^\bullet, \varepsilon)$ be the noisy aggregate flow, and $\hat{\tau} = \tau^*(\hat{\mathbf{y}})$ be a vector of constant tolls. Then with probability at least $1-\beta$ for $\beta > 0$, there* *are at most $\sqrt{n\alpha/4m\gamma}$ players who are $\hat{\zeta}_\varepsilon$-unsatisfied players in $\Gamma^{\hat{\tau}}$ with respect to the congestion $\hat{\mathbf{y}}$, for*

$$\hat{\zeta}_\varepsilon = 4\sqrt{mn\gamma\alpha} + 8\frac{\gamma m^2 \log(m/\beta)}{\varepsilon}. \tag{9}$$

We will now give a rough sketch of the proof, . The full proof appears in Appendix B.

Proof Sketch. First, we will consider the routing game Γ^{τ^*} under the (functional) marginal-cost toll. We will also assume for now that we have the exact congestion $\mathbf{y}^\bullet = \sum_i \mathbf{x}_i^\bullet$. Recall from Lemma 3.2 that the potential function Ψ for this game is equal to the total congestion cost $n \cdot \phi$. Since $\mathbf{x}^\bullet$ is an α-approximate optimal flow, it also approximately minimizes Ψ up to error $n \cdot \alpha$. The construction of Ψ is such that if a player who is ρ-unsatisfied with respect to $\mathbf{y}^\bullet$ plays her best response, then Ψ decreases by at least ρ. Therefore the number of ρ-unsatisfied players with respect to $\mathbf{y}^\bullet$ is at most $n\alpha/\rho$. Here we are intentionally being slightly imprecise to ease exposition. See the full proof for details.

Now, consider the routing game $\Gamma^\tau = (G, \ell + \tau, \mathbf{s})$ that arises from using the constant tolls $\tau = \tau^*(\mathbf{y}^\bullet)$. Note that under functional tolls τ^*, when a player best responds, the tolls may change, however under constant tolls τ the tolls do not change. This might increase the number of players who can gain by deviating. However, notice that when one player changes their route, the tools τ_e^* and τ_e can only change by γ, since τ_e^* is γ-Lipschitz. Thus changing from tolls τ^* to τ can only change the cost any player faces on any route by $m\gamma$. Therefore, we can argue that the number of $(\rho + 2m\gamma)$-unsatisfied players with respect to $\mathbf{y}^\bullet$ in the game Γ^τ is also at most $n\alpha/\rho$.

The last issue to address is that we compute the tolls from the noisy congestion $\hat{\mathbf{y}}$ instead of the exact congestion $\mathbf{y}^\bullet$. This has two effects: 1) the constant tolls $\hat{\tau} = \tau^*(\hat{\mathbf{y}})$ are different from the constant tolls $\tau = \tau^*(\mathbf{y}^\bullet)$ analyzed above and 2) we want to measure the number of unsatisfied players with respect to $\hat{\mathbf{y}}$ instead of $\mathbf{y}^\bullet$. We can address both of these issues using the fact that the noise is small on every edge. Therefore $|y_e - \hat{y}_e|$ is small, and since τ_e^* is Lipschitz, $|\tau_e - \hat{\tau}_e|$ is small as well. In the full proof we carefully account for the magnitude of the noise and its effect on the cost faced by each player to obtain the guarantees stated in the lemma. □

We have so far shown that there might be a few players that are unsatisfied with their current route in $\Gamma^{\hat{\tau}} = (G, \ell + \hat{\tau}, \mathbf{s})$ when they only know a perturbed version of the congestion $\hat{\mathbf{y}}$. We then let these unsatisfied players simultaneously change routes to the routes with the lowest cost (according to the cost $c_{\mathbf{x}_i}(\mathbf{y})$). This procedure, `P-BR`, is detailed in Algorithm 3.

We are now ready to show that the final flow assignments $\hat{\mathbf{x}}$ resulting from the procedure `P-BR`$(\Gamma^{\hat{\tau}}, \mathbf{x}^\bullet, \hat{\zeta}_\varepsilon)$, where $\mathbf{x}^\bullet$ is an α-approximate optimal flow in Γ and $\hat{\zeta}_\varepsilon$ is given in (9), forms an approximate Nash equilibrium in the game $\Gamma^{\hat{\tau}}$ and remains an approximately optimal flow for the original routing game instance Γ.

Lemma 3.7. *Fix any $\alpha > 0$ and $\beta, \varepsilon \in (0,1)$. Let $\Gamma = (G, \ell, \mathbf{s})$ be a routing game and $\mathbf{x}^\bullet$ be an α-approximately optimal flow in Γ. Let $\hat{\mathbf{x}} =$ `P-BR`$(\Gamma^{\hat{\tau}}, \hat{\mathbf{y}}, \mathbf{x}^\bullet, \hat{\zeta}_\varepsilon)$ for $\hat{\zeta}_\varepsilon$ given in (9), $\hat{\mathbf{y}} =$ `P-CON`$(\mathbf{x}^\bullet, \varepsilon)$, and $\hat{\tau} = \tau^*(\hat{\mathbf{y}})$. Then with probability at least $1 - \beta$, $\hat{\mathbf{x}}$ is an $\eta_{eq}(\alpha)$-Nash flow in $\Gamma^{\hat{\tau}} = (G, \ell + \hat{\tau}, \mathbf{s})$ where*

$$\eta_{eq}(\alpha) = O\left(\sqrt{mn\alpha} + \frac{m^2 \log(m/\beta)}{\varepsilon}\right). \tag{10}$$

and $\hat{\mathbf{x}}$ is an $\eta_{opt}(\alpha)$-approximate Nash flow in Γ where

$$\eta_{opt}(\alpha) = O\left(\alpha + \sqrt{mn\alpha}\right). \tag{11}$$

Algorithm 3 Private Best Responses

```
Input: Routing game instance $\Gamma$, congestion $\mathbf{y}$, flow assignment $\mathbf{x}$, satisfaction parameter $\zeta$
Output: New flow assignment $\hat{\mathbf{x}}$
  procedure P-BR($\Gamma, \mathbf{y}, \mathbf{x}, \zeta$)
      Let $\hat{\mathbf{x}} \leftarrow \mathbf{x}$
      for each player $i \in [n]$ do
          if $i$ with flow $\mathbf{x}_i$ is $\zeta$-unsatisfied with respect to congestion $\mathbf{y}$ in game $\Gamma$ then
              Replace $\hat{\mathbf{x}}_i$ by the route with the lowest cost given congestion $\mathbf{y}$.
                  $\hat{\mathbf{x}}_i \leftarrow \operatorname*{argmax}_{\mathbf{x}'_i} \left\{ c_{\mathbf{x}'_i}\left(\mathbf{y}'\right) \right\}$    (breaking ties arbitrarily)
              Where $y'_e = y_e - 1$ if $x_{i,e} = 1, x'_{i,e} = 0; y'_e = y_e + 1$ if $x_{i,e} = 0, x'_{i,e} = 1$; else $y_e = y'_e$.
      return $\hat{\mathbf{x}}$
  end procedure
```

Proof. First, to show that $\hat{\mathbf{x}}$ forms an approximate Nash flow, we need to argue that all players are approximately satisfied with respect to the actual congestion $\mathbf{y} = \sum_i \hat{\mathbf{x}}_i$. As an intermediate step, we will first show that all players in $\hat{\mathbf{x}}$ are approximately satisfied with the input perturbed congestion $\hat{\mathbf{y}}$.

By Lemma 3.6, we know that the number of $\hat{\zeta}_\varepsilon$-unsatisfied players that deviate in our instantiation of P-BR is bounded by

$$\sqrt{n\alpha}/(2\sqrt{m\gamma}) \equiv K.$$

After these players' joint deviation, the congestion on any path is changed by at most $m\,K$, so the total cost on any path is changed by at most $m\gamma K = \sqrt{mn\alpha\gamma}/2$. Therefore, the players that deviate are $\sqrt{mn\alpha\gamma}$-satisfied in $\Gamma^{\hat{\tau}}$ with respect to congestion $\hat{\mathbf{y}}$ after the simultaneous moves. Similarly, the players that were originally $\hat{\zeta}_\varepsilon$-satisfied in $\Gamma^{\hat{\tau}}$ with congestion $\hat{\mathbf{y}}$ remain $(\hat{\zeta}_\varepsilon + \sqrt{mn\alpha\gamma})$-satisfied with $\hat{\mathbf{y}}$ even after the joint deviations.

From standard bounds on the tails of Laplace distribution(Lemma A.4), we can bound the difference between $\hat{\mathbf{y}}$ and $\sum_i \mathbf{x}^\bullet_i$: with probability at least $1 - \beta$,

$$\|\hat{\mathbf{y}} - \sum_i \mathbf{x}^\bullet_i\|_\infty \le 2m\log(m/\beta)/\varepsilon$$

Since the number of players that deviate in P-BR is bounded by K, we could bound $\|\mathbf{y} - \sum_i \mathbf{x}^\bullet_i\|_\infty \le K$. By triangle inequality, we get

$$\|\hat{\mathbf{y}} - \mathbf{y}\|_\infty \le 2m\log(m/\beta)/\varepsilon + K.$$

Since all players in $\hat{\mathbf{x}}$ are $(\hat{\zeta}_\varepsilon + \sqrt{mn\alpha\gamma})$-satisfied with congestion $\hat{\mathbf{y}}$, by Lemma B.4, we knowthat they are also η_{eq}-satisfied with the actual congestion $\mathbf{y}$, where

$$\eta_{eq} = \hat{\zeta}_\varepsilon + \sqrt{mn\alpha\gamma} + \frac{4\gamma m^2\log(m/\beta)}{\varepsilon} + 2K\gamma m = 6\sqrt{mn\alpha\gamma} + \frac{12\gamma m^2\log(m/\beta)}{\varepsilon}.$$

Hence, the flow $\hat{\mathbf{x}}$ forms an η_{eq}-approximate Nash flow in game $\Gamma^{\hat{\tau}}$. To bound the cost of $\hat{\mathbf{x}}$, note that for each edge e, the number of players can increase by at most K. Let $\mathbf{y}^\bullet = \sum_i \mathbf{x}^\bullet_i$, then for each edge, $y_e\ell_e(y_e) - y^\bullet_e\ell_e\left(y^\bullet_e\right) \le n\gamma K + n\gamma = nK(\gamma + 1)$.

Therefore, the average cost for $\hat{\mathbf{x}}$ is

$$\phi(\hat{\mathbf{x}}) = \frac{1}{n}\sum_{e\in E} y_e\ell_e(y_e) \le \frac{1}{n}\sum_{e\in E} y_e^\bullet \ell_e(y_e^\bullet) + mK(\gamma+1) \le \mathrm{OPT}(\mathbf{s}) + \alpha + \frac{\sqrt{mn\gamma\alpha}}{2} + \frac{\sqrt{mn\alpha}}{2\sqrt{\gamma}}$$

This completes the proof. □

3.3 Analysis of FlowToll

Now that we have analyzed the subroutines P-CON and P-BR along with computing the private tolls $\hat{\tau}$, we are ready to analyze the complete mediator FlowToll. Note that in this analysis we will assume that the subroutine P-GD is a blackbox that is JDP and computes an approximately optimal flow in Γ.

We first prove that the mediator FlowToll is JDP. This will give the first condition we require of our mediator in Theorem 2.7. A useful tool in proving mechanisms are JDP is the billboard lemma, which states at a high level that if amechanism can be viewed as posting some public signal (i.e. as if on "a billboard") that is DP in the players' demands, from which (together with knowledge of their own demand) players can derive their part of the output of the mechanism, then the resulting mechanism is JDP.

Lemma 3.8 (Billboard Lemma Rogers and Roth (2014); Hsu et al. (2014a)). *Let $\mathcal{M} : \mathcal{S}^n \to \mathcal{O}$ be an (ε, δ)-differentially private mechanism and consider any function $\theta : \mathcal{S} \times \mathcal{O} \to \mathcal{A}$. Define the mechanism $M' : \mathcal{S}^n \to \mathcal{A}^n$ as follows: on input $\mathbf{s}$, $\mathcal{M}'$ computes $o = \mathcal{M}(\mathbf{s})$, and then $\mathcal{M}'(\mathbf{s})$ outputs to each i:*

$$\mathcal{M}'(\mathbf{s})_i = \theta(s_i, o).$$

$\mathcal{M}'$ is then (ε, δ)-jointly differentially private.

We show that FlowToll is jointly differentially private via the billboard lemma.

Theorem 3.9. *For $\varepsilon, \delta, \beta > 0$, the procedure FlowToll$(\Gamma, \varepsilon, \delta, \beta)$ in Algorithm 1 is (ε, δ)-joint differentially private in the player's input demands $\mathbf{s}$.*

Proof. In order to show JDP using the Billboard Lemma, we need to show that for each player i, the output flow $\hat{\mathbf{x}}_i$ and toll vector $\hat{\tau}$ can be computed only based on i's demands s_i and some (ε, δ)-DP signal.

In Theorem 4.2, we show that the subroutine P-GD$(\Gamma, \varepsilon/4, \delta/2, \beta/2)$ operates in the Billboard model, and can be computed from some $(\varepsilon/4, \delta/2)$-DP billboard signal Λ.

Note that the output flow $\hat{\mathbf{x}}_i$ for each player i produced by P-BR$(\Gamma^{\hat{\tau}}, \hat{\mathbf{y}}, \mathbf{x}^\bullet, \hat{\zeta}_{\varepsilon/4})$ is just a function of the perturbed congestion $\hat{\mathbf{y}}$, $\mathbf{x}_i^\bullet$ and player i's demand. Recall that we know that $\hat{\mathbf{y}} =$ P-CON$(\mathbf{x}^\bullet, \varepsilon/4)$ is $(3\varepsilon/4, \delta/2)$-DP in $\mathbf{s}$ by Corollary 3.4. Therefore, the output flow $\hat{\mathbf{x}}_i$ for each i is just a function of the (ε, δ)-DP signal $(\Lambda, \hat{\mathbf{y}})$, and i's demand s_i. Also, the tolls $\hat{\tau}$ are computed as a function only of $\hat{\mathbf{y}}$. Therefore, by the Billboard Lemma 3.8, the mediator FlowToll$(\Gamma, \varepsilon, \delta, \beta)$ satisfies (ε, δ)-JDP. □

Now we give the appropriate choices of the parameters $(\varepsilon, \delta, \beta)$ for FlowToll$(\Gamma, \varepsilon, \delta, \beta)$ that leads to our main result in the following theorem. This result follows from instantiating Theorem 2.7 with a JDP algorithm that computes an approximately optimal flow $\hat{\mathbf{x}}$ and tolls $\hat{\tau}$ such that $\hat{\mathbf{x}}$ forms an approximate equilibrium in the routing game with tolls $\hat{\tau}$.

Proof of Theorem 2.8. Given any routing game instance $\Gamma = (G, \ell, \mathbf{s})$, we first show that `FlowToll` is a mediator that makes good behavior an η-approximate Nash equilibrium of the mediated game $\Gamma_{\texttt{FlowToll}}$ for $\eta = \tilde{O}\left(m^{3/2} n^{4/5}\right)$.

We assume that $\texttt{P-GD}(\Gamma, \varepsilon/2, \delta/2, \beta/2)$ produces an α-approximate optimal flow $\mathbf{x}^\bullet$ with probability $1 - \beta/2$ (leaving the formal proofs to Theorem 4.2 and Theorem 4.10) where α is given in (5). Consider the instantiation of $\texttt{FlowToll}(\Gamma, \varepsilon, \delta, \beta)$ with

$$\varepsilon = \frac{\sqrt{m}}{n^{1/5}}, \quad \delta = n^{-2}, \quad \beta = n^{-2}.$$

Given the functional tolls τ^* defined in (6) and the fact that if we ever get an edge congestion $\hat{y}_e > n$ from the output of `P-CON` then we round it down to n, so the edge tolls $\hat{\tau}_e$ are never bigger than $n\gamma$. Using our bound for $\eta_{eq}(\alpha)$ in (10) and setting $\eta_{eq} = \eta_{eq}(\alpha)$ where α is as above, from Theorem 2.7 we have with probability $1-\beta$ the bound $\eta \le \eta_{eq} + m(U+n)(2\varepsilon+\beta+\delta) = \tilde{O}\left(m^{3/2} n^{4/5}\right)$.

We then show that good behavior results in an η_{opt}-approximately optimal flow for the original routing game instance Γ, where

$$\eta_{opt} = \tilde{O}\left(m n^{4/5}\right).$$

It then follows that $\eta_{opt} = \eta_{opt}(\alpha)$ from (11) and for α given in (5). □

4 Computing an Approximately Optimal Flow under JDP

In this section we show how to compute an approximately optimal flow $\mathbf{x}^\bullet$ under joint differential privacy. We first consider a convex relaxation of the problem of minimizing social cost in the routing game instance $(\Gamma, \ell, \mathbf{s})$. Let $\mathcal{F}^R(\mathbf{s}) \subseteq [0,1]^{n \times m}$ be the set of feasible *fractional flows* (i.e. the convex relaxation of the set $\mathcal{F}(\mathbf{s})$). Then the optimal fractional flow is given by the convex program:

$$\begin{aligned} \min \quad & c(\mathbf{y}) = \frac{1}{n} \sum_{e \in E} y_e \ell_e(y_e) && (12) \\ \text{such that} \quad & \mathbf{x} \in \mathcal{F}^R(\mathbf{s}) \subseteq [0,1]^{n \times m} \\ & y_e = \sum_{i=1}^{n} x_{i,e} \qquad \forall e \in E, \quad \forall i \in [n] && (13) \end{aligned}$$

Note that the second derivative of $y_e \ell_e(y)$ is $2\ell'_e(y_e) + y_e \ell''_e(y_e)$. Since ℓ_e is assumed to be convex and nondecreasing, the second derivative is non-negative as long as $y_e \ge 0$. Hence the objective function c of this program is indeed convex on the feasible region.

We write $\mathcal{G}^R(\mathbf{s}) := \mathcal{F}^R(\mathbf{s}) \times [0,n]^m$ to denote the space where the decision variables reside, i.e. $(\mathbf{x}, \mathbf{y}) \in \mathcal{G}^R(\mathbf{s})$. Given any demands $\mathbf{s}$, we write $\mathrm{OPT}^R(\mathbf{s})$ to denote the optimal objective value of the convex program and $\mathrm{OPT}(\mathbf{s})$ to be the optimal objective value when $\mathbf{x} \in \mathcal{F}(\mathbf{s})$. Note that we always have $\mathrm{OPT}^R(\mathbf{s}) \le \mathrm{OPT}(\mathbf{s})$

Our goal is to first compute an approximately optimal solution to the relaxed convex program, and then round the resulting fractional solution to be integral. We then show that the final solution is an approximately optimal flow to the original instance Γ.

4.1 The JDP Gradient Descent Algorithm

We will work extensively with the Lagrangian of our problem. For each constraint of (13), we introduce a dual variable λ_e. The Lagrangian is then

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) = c(\mathbf{y}) - \sum_{e \in E} \lambda_e \left(\sum_i x_{i,e} - y_e \right).$$

Since our convex program satisfies Slater's condition Slater (1959), we know that strong duality holds:

$$\max_{\lambda \in \mathbb{R}^m} \min_{(\mathbf{x},\mathbf{y}) \in \mathcal{G}^R(\mathbf{s})} \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) = \min_{(\mathbf{x},\mathbf{y}) \in \mathcal{G}^R(\mathbf{s})} \max_{\lambda \in \mathbb{R}^m} \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) = \mathrm{OPT}^R(\mathbf{s}). \tag{14}$$

We will interpret the Lagrangian objective as the payoff function of a zero-sum game between the minimization player, who plays flows $\mathbf{z} = (\mathbf{x}, \mathbf{y})$, and the maximization player, who plays dual variables λ. We will abuse notation and write $\mathcal{L}(\mathbf{z}, \lambda) = \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda)$. We refer to the game defined by this payoff matrix the *Lagrangian game*. We will privately compute an approximate equilibrium of the Lagrangian game by simulating repeated plays between the two players. In each step, the dual player will play an approximate best response to the flow player's strategy. The flow player will update his flow using a no-regret algorithm.

In particular, the flow player uses an online gradient descent algorithm to produce a sequence of T actions $\{\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(T)}\}$ based on the loss functions given by the dual player's actions $\{\lambda^{(1)}, \dots, \lambda^{(T)}\}$. At each round $t = 1, \dots, T$, the flow player will update both $\mathbf{x}^{(t)}$ and $\mathbf{y}^{(t)}$ using the projected gradient update step GD in Algorithm 4.

```
Algorithm 4 Gradient Descent with Projection
Input: Convex feasible domain $\mathcal{D}$, a convex function $r$, some $\omega \in \mathcal{D}$, and learning parameter $\eta$.
Output: Some new $\omega' \in \mathcal{D}$.
  procedure GD($\mathcal{D}, r, \omega, \eta$)
      We define the projection map $\Pi_{\mathcal{D}}$ as
          $\Pi_{\mathcal{D}}(v') = \operatorname{argmin}_{v \in \mathcal{D}} ||v - v'||_2$
      We then set
          $\omega' \leftarrow \Pi_{\mathcal{D}}(\omega - \eta \nabla r(\omega))$
  return $\omega'$
  end procedure
```

In order to reason about how quickly the projected gradient procedure converges to an approximately optimal flow, we need to bound the diameter of the space of dual solutions. We will also need to argue that bounding the space of feasible dual solutions does not affect the value of the game. Specifically, we will bound the dual players' action to the set

$$\mathcal{B} = \{\lambda \in \mathbb{R}^m \mid \|\lambda\|_1 \leq 2m\}, \tag{15}$$

Then fixing a flow played by the primal player, the dual player's best response is simply to select an edge e where the constraint (13) is most violated and set $\lambda_e^{(t)} = \pm 2m$. Notice that, since the constraints depend on the source/sink pairs, and we need to ensure joint differential privacy with respect to this data, we must select the most violated constraint in a way that maintains privacy. Using a straightforward application of the DP exponential mechanism McSherry and Talwar (2007), we can obtain a constraint that is approximately the most violated. Since this step is standard, we defer the details to the appendix.

From the repeated plays of the Lagrangian Game, we will obtain a fractional solution $\overline{\mathbf{z}} = (\overline{\mathbf{x}}, \overline{\mathbf{y}})$ to the convex program. Finally, we will round the fractional flow $\overline{\mathbf{x}}$ to an integral solution $\mathbf{x}^\bullet$ for the original minimum-cost flow instance $\Gamma = (G, \ell, \mathbf{s})$ using the rounding procedure PSRR proposed by Raghavan and Thompson (1987), given in Algorithm 5. The full procedure P-GD is given in Algorithm 6.

Algorithm 5 Path Stripping and Randomized Rounding

```
Input: A fractional flow solution $\overline{\mathbf{x}}_i \in \mathcal{F}^R(\mathbf{s}_i)$ for player $i$
Output: An integral flow solution $\mathbf{x}_i \in \mathcal{F}(\mathbf{s}_i)$ for player $i$
  procedure PSRR($\overline{\mathbf{x}}_i$)
      Let $\Lambda_i = \{P_j\}$ be the set of $(s_i^1, s_i^2)$-paths in $G$
      for each path $P_j$ do
          Let $w_j = \min\{\overline{x}_{i,e} \mid e \in P_j\}$
          for each edge $e \in P_j$ do
              Let $\overline{x}_{i,e} \leftarrow x_{i,e} - w_j$
      Sample a path $P$ from $\Lambda_i$ such that $\mathbb{P}[P = P_j] = w_j$
      for each edge $e \in E$ do
          Let $x_{i,e} = \mathbb{I}[e \in P]$
      return $\mathbf{x}_i$
  end procedure
```

4.2 Privacy of the JDP Gradient Descent Algorithm

We will use Lemma 3.8 (the *billboard lemma*) to prove that P-GD satisfies joint differential privacy. We first show that the sequence of plays by the dual player satisfies standard differential privacy.

Lemma 4.1. *The sequence $\{\lambda^{(t)}\}_{t=1}^T$ in P-GD$(\Gamma, \varepsilon, \delta, \beta)$ satisfies (ε, δ)-differential privacy in the reported types $\mathbf{s}$ of the players.*

Proof. At each iteration of the main for-loop, we use the exponential mechanism with quality score q to find which edge e has the most violated constraint in (13). By Lemma A.5, each tuple[3] $(\bullet^{(t)}, e^{(t)})$is ε'-differentially private. Note that the dual strategy $\lambda^{(t)}$ is simply a post-processing function of the tuple $(\bullet^{(t)}, e^{(t)})$, and by Lemma A.1, we know that $\lambda^{(t)}$ is ε'-differentially private. By the composition theorem for differential privacy (Lemma A.8), we know that the sequence of the dual plays $\lambda^{(t)}$ satisfies (ε, δ)-differential privacy, with the assignment of ε' in P-GD.

□

We are now ready to show that our algorithm satisfies joint differential privacy.

Theorem 4.2. *P-GD$(\Gamma, \varepsilon, \delta, \beta)$ given in Algorithm 6 is (ε, δ)-jointly differentially private.*

Proof. In order to establish joint differential privacy using the Billboard Lemma (Lemma 3.8), we just need to show that the output solution $\{\mathbf{x}_i\}$ for each player i is just a function of the dual plays $\{\lambda^{(t)}\}$ and i's private data.

[3] Recall that $\bullet \in \{+, -\}$ indicating whether $\sum_i x_{i,e} > y_e$ or $\sum_i x_{i,e} < y_e$

Algorithm 6 Computing Approximately Optimal Flow via JDP Gradient Descent

```
Input: Routing Game $\Gamma = (G, \ell, \mathbf{s})$; privacy parameters $(\varepsilon, \delta)$; failure probability $\beta$
Output: $\mathbf{x}_i^\bullet$, a $s_i = (s_i^1, s_i^2)$ flow for each player $i \in [n]$
  procedure P-GD$(\Gamma, \epsilon, \delta, \beta)$
      Define the following quantities:

      $T \leftarrow \Theta\left(\frac{\epsilon n\sqrt{m}}{\log(mn/\beta)\sqrt{\log(1/\delta)}}\right) \qquad \varepsilon' \leftarrow \varepsilon/\sqrt{8T\ln(1/\delta)} \qquad \eta_y \leftarrow \frac{D_y}{G_y\sqrt{T}} \qquad \eta_x \leftarrow \frac{D_x}{G_x\sqrt{T}}$

      $G_y \leftarrow \sqrt{(m-1)(\gamma+1)^2 + (\gamma+1+2m)^2} \qquad D_y \leftarrow n\sqrt{m}$
      $G_x \leftarrow 2m\sqrt{n} \qquad D_x \leftarrow \sqrt{mn}$

      Initialize: $\mathbf{y}^{(1)} \in [0, n]^m$ and $\mathbf{x}^{(1)} \in \mathcal{F}^R(\mathbf{s})$. Let $\mathbf{z}^{(1)} \leftarrow (\mathbf{x}^{(1)}, \mathbf{y}^{(1)})$
      Define the quality score $q : \mathcal{G}(\mathbf{s}) \times ((+,-) \times E) \to \mathbb{R}$:

      $f_e(\mathbf{z}) \leftarrow \sum_i x_e^i - y_e \qquad q(\mathbf{z}, (+, e)) \leftarrow +f_e(\mathbf{z}) \qquad q(\mathbf{z}, (-, e)) \leftarrow -f_e(\mathbf{z}).$

      for $t = 1, \cdots, T$ do
          Let $(\bullet^{(t)}, e^{(t)}) \leftarrow \mathcal{M}_E(\mathbf{s}, q, \epsilon')$ (The Exponential Mechanism)
          Approximate best-response for the dual player $\lambda^{(t)}$:
          if $\bullet^{(t)} = +$   then   $\lambda_{e^{(t)}}^{(t)} \leftarrow -2m$
          else   $\lambda_{e^{(t)}}^{(t)} \leftarrow +2m$
          for $e' \in E \setminus \{e^{(t)}\}$   do   $\lambda_{e'}^{(t)} = 0$
          Gradient descent update on the primal:
          Take a step to improve the individual flow variables $\mathbf{x}^{(t)}$:

              $\mathbf{x}^{(t+1)} \leftarrow \mathtt{GD}(\mathcal{F}(\mathbf{s}), \mathcal{L}(\cdot, \mathbf{y}^{(t)}, \lambda^{(t)}), \mathbf{x}^{(t)}, \eta_x)$

          Take a step to improve the congestion variables $\mathbf{y}^{(t)}$:

              $\mathbf{y}^{(t+1)} \leftarrow \mathtt{GD}([0, n]^m, \mathcal{L}(\mathbf{x}^{(t)}, \cdot, \lambda^{(t)}), \mathbf{y}^{(t)}, \eta_y)$

          Let $\mathbf{z}^{(t+1)} = (\mathbf{x}^{(t+1)}, \mathbf{y}^{(t+1)})$ be the new action for the primal player.
      $\overline{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}^t$ and $\overline{\lambda} = \frac{1}{T}\sum_{t=1}^{T} \lambda^{(t)}$
      for each player $i$ do round the fractional flow: $\mathbf{x}_i^\bullet \leftarrow \mathtt{PSRR}(\overline{\mathbf{x}}_i)$
      return $\mathbf{x}^\bullet = (\mathbf{x}_i^\bullet)_{i\in[n]}$
  end procedure
```

Note that initially, each player i simply sets $\mathbf{x}_i^{(1)}$ to be a feasible flow in the set $\mathcal{F}^R(\mathbf{s}_i)$, which only depends on i's private data.

Then at each round t, the algorithm updates the vector $\mathbf{x}^{(t)}$ using the gradient:

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mathbf{y}^{(t)}, \lambda^{(t)}) = \left(\left(-\lambda_e^{(t)}\right)_{e\in E}\right)_{i\in[n]}.$$

The gradient descent update for $\mathbf{x}^{(t)}$ is

$$\begin{aligned}\mathbf{x}^{(t+1)} &= \Pi_{\mathcal{F}^R(\mathbf{s})}\left[\mathbf{x}^{(t)} - \eta_x(\lambda^{(t)})_{i\in[n]}\right]\\ &= \arg\min_{\mathbf{x}\in\mathcal{F}^R(\mathbf{s})}\left\|\mathbf{x} - \left(\mathbf{x}^{(t)} - \eta_x(\lambda^{(t)})_{i\in[n]}\right)\right\|_2^2\\ &= \arg\min_{\mathbf{x}\in\mathcal{F}^R(\mathbf{s})}\sum_{i\in[n]}\left(\sum_{e\in E}\left\|x_{i,e} - (x_{i,e}^{(t)} - \eta_x\lambda_e^{(t)})\right\|_2^2\right)\end{aligned}$$

Note that this update step can be decomposed into n individual updates over the players:

$$\mathbf{x}_i^{(t+1)} = \arg\min_{\mathbf{x}\in\mathcal{F}^R(\mathbf{s}_i)}\sum_{e\in E}\left\|x_{i,e} - \left(x_{i,e}^{(t)} - \eta_x\lambda_e^{(t)}\right)\right\|_2^2$$

Since such an update only depends on the private data of i and also the sequence of dual plays $\{\lambda^{(t)}\}$, we know that $\{\mathbf{x}^{(t)}\}$ satisfies (ε, δ)-joint differential privacy by the Billboard Lemma.

Finally, the output integral solution $\mathbf{x}_i$ to each player i is simply a sample from the distribution induced by the average play of i: $\overline{\mathbf{x}}$. Therefore, we can conclude that releasing $\mathbf{x}_i$ to each player i satisfies (ε, δ)-joint differential privacy.

□

4.3 Utility of the JDP Gradient Descent Algorithm

We now establish the accuracy guarantee of the integral flow $\mathbf{x}^\bullet$ computed by the procedure P-GD. First, consider the average of the actions taken by both players over the T rounds of the algorithm P-GD: $\overline{\mathbf{z}} = \frac{1}{T}\sum_{t=1}^T \mathbf{z}^{(t)}$ and $\overline{\lambda} = \frac{1}{T}\sum_{t=1}^T \lambda^{(t)}$. Recall that the *minimax value* of the Lagrangian game is defined as

$$\max_{\lambda\in\mathbb{R}^m}\min_{(\mathbf{x},\mathbf{y})\in\mathcal{G}^R(\mathbf{s})}\mathcal{L}(\mathbf{x},\mathbf{y},\lambda) = \min_{(\mathbf{x},\mathbf{y})\in\mathcal{G}^R(\mathbf{s})}\max_{\lambda\in\mathbb{R}^m}\mathcal{L}(\mathbf{x},\mathbf{y},\lambda) = \mathrm{OPT}^R(\mathbf{s}). \tag{16}$$

Thus, in order to show that $\overline{\mathbf{z}}$ is a flow with nearly optimal cost (i.e. cost not much larger than $\mathrm{OPT}^R(\mathbf{s})$), it suffices to show that $(\overline{\mathbf{z}}, \overline{\lambda})$ are a pair of "approximate minimax strategies". That is, each player is guaranteeing itself a payoff that is close to the value of the game. Formally, $(\overline{\mathbf{z}}, \overline{\lambda})$ is a pair of $\mathcal{R}$-*approximate minimax strategies* if

$$\forall \mathbf{z}',\ \mathcal{L}(\overline{\mathbf{z}}, \overline{\lambda}) \le \mathcal{L}(\mathbf{z}', \overline{\lambda}) + \mathcal{R} \qquad \text{and} \qquad \forall \lambda',\ \mathcal{L}(\overline{\mathbf{z}}, \overline{\lambda}) \ge \mathcal{L}(\overline{\mathbf{z}}, \lambda') - \mathcal{R}.$$

Looking ahead, using the properties of gradient descent, we can show that $(\overline{\mathbf{z}}, \overline{\lambda})$ are a pair of $\mathcal{R}$-approximate minimax strategies for a bound $\mathcal{R}$ that will grow with the norm of the dual player's actions, i.e. $\|\lambda^{(t)}\|_2$. Thus, in P-GD, the dual player's action is chosen to have bounded norm (at most $2m$), in order to ensure $\mathcal{R}$ is relatively small. However, from (16) it's not clear that the optimal dual strategy has small norm, so restricting the norm of the dual player's actions might change the value of the game. However, we show that restricting the norm of the dual player's action does not change the value of the game.

Let $(\mathbf{z}^*, \lambda^*)$ be a pair of (exact) minimax strategies in the Lagrangian game. By strong duality, we know that

$$\mathcal{L}(\mathbf{z}^*, \lambda^*) = \mathrm{OPT}^R(\mathbf{s})$$

and $\mathbf{z}^*$ is an optimal and feasible solution. We now reason about the *restricted Lagrangian game*, in which the dual player's action is restricted to the space $\mathcal{B} = \{\lambda \in \mathbb{R}^m \mid \|\lambda\|_1 \leq 2m\} \subseteq \mathbb{R}^m$. The next lemma states that even when the dual player's actions are restricted, then $\mathbf{z}^*$ is still a minimax strategy for the primal player. That is, the primal player cannot take advantage of the restriction on the dual player to obtain a higher payoff.

Lemma 4.3. *There exists a dual strategy $\lambda^*_{\mathcal{B}} \in \mathcal{B}$ such that $(\mathbf{z}^*, \lambda^*_{\mathcal{B}})$ is a pair of (exact) minimax strategies for the restricted Lagrangian game.*

Proof. Since $\mathbf{z}^*$ is an (exact) minimax strategy for the (unrestricted) Lagrangian game, we know that for any $\lambda \in \mathcal{B}$

$$\mathcal{L}(\mathbf{z}^*, \lambda) = c(\mathbf{y}^*) = \mathrm{OPT}^R(\mathbf{s}).$$

Let $\mathbf{x}' \in \mathcal{F}^R(\mathbf{s})$ and $\mathbf{y} \in [0, n]^m$ be different flows such that $\mathbf{x}' \neq \mathbf{x}^*$ and $\mathbf{y}' \neq \mathbf{y}^*$. We want to show that

$$\max_{\lambda \in \mathcal{B}} \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) \geq \max_{\lambda \in \mathcal{B}} \mathcal{L}(\mathbf{z}^*, \lambda) := \mathcal{L}(\mathbf{z}^*, \lambda^*_{\mathcal{B}}).$$

If we have $y'_e = \sum_{i=1}^n x'_{i,e}$ for all $e \in E$, then

$$\max_{\lambda \in \mathcal{B}} \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) = c(\mathbf{y}) \geq c(\mathbf{y}^*).$$

Suppose there is some edge such that $y'_e \neq \sum_{i=1}^n x'_{i,e}$, then we define $\Delta := \|\mathbf{y}' - \sum_{i=1}^n \mathbf{x}'_i\|_\infty$. With the cost function in terms of the individual flow variables in (4) we know that

$$c(\mathbf{y}) \geq \phi(\mathbf{x}) - \frac{1}{n} \sum_{e \in E} \Delta \cdot \ell_e(n) \geq \phi(\mathbf{x}) - m \cdot \Delta \geq c(\mathbf{y}^*) - m \cdot \Delta.$$

Note that the dual player can set $\lambda_e = 2m$ for $\sum_{i=1}^n x'_{ie} - y'_e > 0$ or $\lambda_e = -2m$ for $\sum_{i=1}^n x'_{ie} - y'_e < 0$ for the maximally violated edge e:

$$\max_{\lambda \in \mathcal{B}} \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) = c(\mathbf{y}) + 2m \cdot \Delta \geq c(\mathbf{y}^*) + \cdot\Delta(2m - m) > \mathrm{OPT}^R(\mathbf{s}).$$

Therefore, any infeasible $(\mathbf{x}, \mathbf{y})$ would suffer loss at least $\mathrm{OPT}^R(\mathbf{s})$ in the worst case over the dual strategy space. It follows that $(\mathbf{z}^*, \lambda^*)$ is a minimax strategy.

Since both players' action spaces $\mathcal{G}^R(\mathbf{s})$ and $\mathcal{B}$ are compact, then there exists a minimax strategy $(\mathbf{z}^*, \lambda^*_{\mathcal{B}})$ of the restricted Lagrangian game. □

Using the previous lemma, we know that the value of the restricted game is the same

Lemma 4.4. *Let $(\mathbf{z}, \lambda)$ be a pair of $\mathcal{R}$-approximate minimax strategy of the restricted Lagrangian game, and $\mathbf{z} = (\mathbf{x}, \mathbf{y})$. Then the fractional solution $\mathbf{x}$ satisfies*

$$\phi(\mathbf{x}) \leq OPT^R(\mathbf{s}) + 4\mathcal{R}.$$

Proof. We will first bound the constraint violation in $(\mathbf{x}, \mathbf{y})$. Let $e' \in \arg\max_{e\in E} |\sum_i x_{i,e} - y_e|$ be an edge where the constraint is violated the most, and let $\Delta = |\sum_i x_{i,e'} - y_{e'}|$. Consider the dual strategy $\lambda' \in \mathcal{B}$ such that

$$\lambda'_{e'} = \begin{cases} -2m & \text{if } \sum_i x_{i,e'} - y_{e'} \geq 0 \\ 2m & \text{otherwise} \end{cases}$$

and $\lambda'_e = 0$ for all $e \neq e'$. Now compare the payoff values $\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda)$ and $\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda')$. By the property of $\mathcal{R}$-approximate equilibrium and letting $((\mathbf{x}^*, \mathbf{y}^*), \lambda^*)$ be the exact equilibrium, we have

$$\begin{aligned} \mathrm{OPT}^R(\mathbf{s}) - \mathcal{R} &= \mathcal{L}(\mathbf{x}^*, \mathbf{y}^*, \lambda^*) - \mathcal{R} \leq \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda^*) - \mathcal{R} \\ &\leq \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) \leq \mathcal{L}(\mathbf{x}^*, \mathbf{y}^*, \lambda) + \mathcal{R} \leq \mathrm{OPT}^R(\mathbf{s}) + \mathcal{R} \\ \Longrightarrow \qquad & \mathrm{OPT}^R(\mathbf{s}) - \mathcal{R} \leq \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) \leq \mathrm{OPT}^R(\mathbf{s}) + \mathcal{R} \end{aligned}$$

and

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda') \leq \mathrm{OPT}^R(\mathbf{s}) + 2\mathcal{R}.$$

Since $(\mathbf{x}, \mathbf{y})$ violates equality constraint on each edge by at most Δ, we know that

$$c(\mathbf{y}) \geq \phi(\mathbf{x}) - \frac{1}{n} \cdot \sum_{e\in E} \left| \sum_i x_{i,e} - y_e \right| \cdot \ell_e(n) \geq \mathrm{OPT}^R(\mathbf{s}) - m\Delta.$$

Also, the penalty incurred by λ' is at least

$$\sum_{e\in E} \lambda'_e \left(y_e - \sum_i x_{i,e} \right) = 2m \cdot \Delta.$$

Therefore, we could bound

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda') \geq \mathrm{OPT}^R(\mathbf{s}) + m \cdot \Delta.$$

It follows that $\Delta \leq 2\mathcal{R}/m$.

Next we will show the accuracy guarantee of $\mathbf{x}$. Consider an all-zero strategy for the dual player λ'', that is $\lambda''_e = 0$ for each $e \in E$. We know such a deviation will not increase the payoff by more than $\mathcal{R}$:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda'') \leq \mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda) + \mathcal{R} \leq \mathrm{OPT}^R(\mathbf{s}) + 2\mathcal{R},$$

and also $\mathcal{L}(\mathbf{x}, \mathbf{y}, \lambda'') = c(\mathbf{y})$, so we must have

$$c(\mathbf{y}) \leq \mathrm{OPT}^R(\mathbf{s}) + 2\mathcal{R}.$$

Now we could give the accuracy guarantee for the cost of the individual flows $\mathbf{x}$:

$$\phi(\mathbf{x}) \leq c(\mathbf{y}) + \frac{1}{n} \sum_{e\in E} \Delta \cdot \ell_e(n) \leq \mathrm{OPT}^R(\mathbf{s}) + 2\mathcal{R} + 2(\mathcal{R}/m) \cdot m = \mathrm{OPT}^R(\mathbf{s}) + 4\mathcal{R}.$$

This completes the proof of the lemma. □

The previous discussion shows that if $(\overline{\mathbf{z}}, \overline{\lambda})$ is a pair of approximate minimax strategies for the restricted Lagrangian game, then $\overline{\mathbf{x}}$ represents an approximately optimal flow. In the remainder of this section, we show that $(\overline{\mathbf{z}}, \overline{\lambda})$ will be such a pair of strategies. To do so, we use a well known result of Freund and Schapire Freund and Schapire (1996), which states that if $\overline{\mathbf{z}}$ and $\overline{\lambda}$ have "low regret," then they are a pair of approximate minimax strategies. "Regret" is defined as follows.

Definition 4.5. Given a sequence of of actions $\{\mathbf{z}^{(t)}\}$ and $\{\lambda^{(t)}\}$ in the Lagrangian game, we define *regret* for each player as:

$$\mathcal{R}_{\mathbf{z}} \equiv \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}(\mathbf{z}^{(t)}, \lambda^{(t)}) - \min_{\mathbf{z}\in\mathcal{G}(\mathbf{s})} \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}(\mathbf{z}, \lambda^{(t)})$$
$$\mathcal{R}_{\lambda} \equiv \max_{\lambda\in\mathcal{B}} \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}(\mathbf{z}^{(t)}, \lambda) - \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}(\mathbf{z}^{(t)}, \lambda^{(t)})$$

Given this definition, the result of Freund and Schapire (1996) can be stated as follows.

Theorem 4.6 (Freund and Schapire (1996))**.** *If $(\overline{\mathbf{z}}, \overline{\lambda})$ is the average of the primal and dual players' actions in P-GD, then $(\overline{\mathbf{z}}, \overline{\lambda})$ is a pair of $(\mathcal{R}_{\mathbf{z}} + \mathcal{R}_{\lambda})$-approximate minimax strategies of the restricted Lagrangian game.*

Given the previous theorem, our goal is now roughly to show that $\overline{\mathbf{z}}$ and $\overline{\lambda}$ have low regret. To do so, we need to analyze the regret properties of the gradient descent procedure, as well as the additional regret incurred by the noise added to ensure joint differential privacy.

Specifically, the gradient descent procedure `GD` satisfies the following regret bound.

Lemma 4.7 (Zinkevich (2003))**.** *Fix the number of steps $T \in \mathbb{N}$. Let $\hat{\mathcal{D}}$ be a convex and closed set with bounded diameter, i.e. for every $\omega, \omega' \in \mathcal{D}$,*

$$||\omega - \omega'||_2 \leq D.$$

Let $r^1, \ldots, r^T$ be a sequence of differentiable, convex functions with bounded gradients, i.e. for every step $t \in [T]$,

$$||\nabla r^t||_2 \leq G.$$

Let $\eta = \frac{D}{G\sqrt{T}}$ and $\omega^0 \in \mathcal{D}$ be arbitrary. Then if we compute $\omega^1, \ldots, \omega^T \in \mathcal{D}$ according to the rule $\omega^{t+1} \leftarrow \mathit{GD}(\mathcal{D}, r^t, \omega^t, \eta^t)$, the sequence $\{\omega^1, \ldots, \omega^T\}$ satisfies

$$R^T(\mathit{GD}) := \sum_{t=1}^{T} r^t(\omega^t) - \min_{\omega\in\mathcal{D}} \left\{ \sum_{t=1}^{T} r^t(\omega) \right\} \leq GD\sqrt{T} \tag{17}$$

We can now use this regret bound for `GD` to give a regret bound for the private gradient descent procedure `P-GD`.

Lemma 4.8. *Fix $\varepsilon, \delta, \beta > 0$. If $(\overline{\mathbf{z}}, \overline{\lambda})$ is the average of the primal and dual players' actions in P-GD$(\Gamma, \varepsilon, \delta, \beta)$, then with probability at least $1 - \beta$, $(\overline{\mathbf{z}}, \overline{\lambda})$ are a pair of $\mathcal{R}$-approximate minimax strategies in the restricted Lagrangian game, for*

$$\mathcal{R} = \tilde{O}\left(\frac{\sqrt{n}m^{5/4}}{\sqrt{\varepsilon}}\right)$$

Proof. In light of Theorem 4.6, we know $\mathcal{R} = \mathcal{R}_z + \mathcal{R}_\lambda$, so we just need to bound the regrets for both players. For the flow player $\mathbf{z}$, we will bound the regrets of $\mathbf{x}$ and $\mathbf{y}$ separately by invoking the regret bound of Zinkevich (2003) given in Lemma 4.7.

We define G_y such that $\forall t \in [T]$ we have $||\nabla_{\mathbf{y}}\mathcal{L}(\mathbf{z}, \lambda^{(t)})||_2 \leq G_y$ and D_y such that $\forall \mathbf{y}, \mathbf{y}' \in [0, n]^m$ we have $||\mathbf{y} - \mathbf{y}'||_2 \leq D_y$. We define corresponding quantities for G_x and D_x. It suffices to set these values in the following way:

$$G_y := \sqrt{(m-1)(\gamma+1)^2 + (\gamma+1+2m)^2} \qquad D_y := n\sqrt{m}$$

$$G_x := 2m\sqrt{n} \qquad D_x := \sqrt{mn}$$

Using (17) we have the following bound on the regret .

$$\begin{aligned}\mathcal{R}_z &\leq 1/\sqrt{T}\left(G_y \cdot D_y + G_x \cdot D_x\right)\\ &\leq \frac{n\sqrt{m}}{\sqrt{T}} \cdot \left(\sqrt{m(\gamma+1)^2 + (\gamma+1+2m)^2} + 2m\right)\\ &= O\left(\frac{nm^{3/2}}{\sqrt{T}}\right) \qquad (18)\end{aligned}$$

with the following step sizes:

$$\eta_x := \frac{D_x}{G_x\sqrt{T}} \qquad \eta_y := \frac{D_y}{G_y\sqrt{T}} \qquad (19)$$

Now we bound the regret for the dual player. Note that each agent could only affect the quality score q of each edge by 1. By the utility guarantee of the exponential mechanism stated in Lemma A.6 we know that with probability at least $1-\beta/T$, at round t

$$\max_{(\bullet,e)\in\{\pm\}\times E}\left|q(\mathbf{z}^{(t)},(\bullet,e)) - q(\mathbf{z}^{(t)},(\bullet^{(t)},e^{(t)})\right| \leq \frac{2(\log(2mT/\beta))}{\varepsilon'} \qquad (20)$$

We condition on this level of accuracy for each round t, which holds except with probability β.

Also, at each round t, a best response for the dual player is to put weight $\pm 2m$ on the edge with the most violation, so we can bound the regret:

$$\begin{aligned}\mathcal{R}_\lambda &= \frac{1}{T}\sum_{t=1}^{T}\left[\max_{\lambda\in\mathcal{B}}\mathcal{L}(\mathbf{z}^{(t)},\lambda) - \mathcal{L}(\mathbf{z}^{(t)},\lambda^{(t)})\right]\\ &\leq \frac{1}{T}\sum_{t=1}^{T} 2m \cdot \frac{2(\log(2mT/\beta))}{\varepsilon'}\\ &= 2m \cdot \frac{2(\log(2mT/\beta))}{\varepsilon'}\end{aligned}$$

For $T = \Theta\left(\frac{\varepsilon n\sqrt{m}}{\log(mn/\beta)\sqrt{\log(1/\delta)}}\right)$, we know that

$$\begin{aligned}\mathcal{R} = \mathcal{R}_z + \mathcal{R}_\lambda &= O\left(\frac{nm^{3/2}}{\sqrt{T}} + \frac{m\log(mT/\beta)\sqrt{T\log(1/\delta)}}{\epsilon}\right)\\ &= O\left(\frac{\sqrt{n}m^{5/4}}{\sqrt{\varepsilon}} \cdot \mathrm{polylog}(1/\delta, 1/\beta, n, m)\right)\end{aligned}$$

□

The previous lemma shows that the fractional solution has nearly optimal cost. The last thing we need to do is derive a bound on how much the rounding procedure PSRR increases the cost of the final integral solution.

Lemma 4.9. *Let $\overline{\mathbf{x}}$ be any feasible fractional solution to the convex program (12), and let $\mathbf{x}^\bullet$ be an integral solution obtained by the rounding procedure PSRR($\overline{\mathbf{x}}$). Then, with probability at least $1-\beta$,*

$$\phi(\mathbf{x}^\bullet) \leq \phi(\overline{\mathbf{x}}) + m(\gamma+1)\sqrt{2n\ln(m/\beta)}.$$

Proof. From the analysis of Raghavan and Thompson (1987) (in Theorem 3.1 of the source), we know that with probability at least $1 - \beta$,

$$\max_{e \in E} \left[\sum_i x^{\bullet}_{i,e} - \sum_i \overline{x}_{i,e} \right] < \sqrt{2n \ln(m/\beta)} \equiv W.$$

Finally, we could bound the difference between the costs $\phi(\mathbf{x}^{\bullet})$ and $\phi(\overline{\mathbf{x}})$

$$\begin{aligned} \phi(\mathbf{x}^{\bullet}) - \phi(\overline{\mathbf{x}}) &\leq \frac{1}{n} \cdot \left[\sum_e x^{\bullet}_{i,e} \cdot \left(\ell_e \left(\sum_j x^{\bullet}_{j,e} \right) - \ell_e \left(\sum_j \overline{x}_{j,e} \right) \right) + \sum_e W \cdot \ell_e \left(\sum_j x^{\bullet}_{j,e} \right) \right] \\ &\leq \frac{1}{n} \cdot (mn\gamma W + mnW) = mW(\gamma + 1). \end{aligned}$$

This completes the proof. □

Combining Lemma 4.8 and Lemma 4.9 we obtain our desired bound on the quality of the joint differentially private integral solution.

Theorem 4.10. *Let $\Gamma = (G, \ell, \mathbf{s})$ be a routing game and $\varepsilon, \delta, \beta > 0$ be parameters. If $\mathbf{x}^{\bullet}$ is the final integral solution given by* `P-GD`$(\Gamma, \varepsilon, \delta, \beta)$*, then with probability at least $1 - \beta$, the cost of $\mathbf{x}^{\bullet}$ satisfies*

$$\phi(\mathbf{x}^{\bullet}) \leq OPT(\mathbf{s}) + \tilde{O}\left(\frac{\sqrt{n} m^{5/4}}{\sqrt{\varepsilon}} + m\sqrt{n} \right)$$

i.e. $\mathbf{x}^{\bullet}$ is an α-approximate optimal flow for $\alpha = \tilde{O}\left(\frac{\sqrt{n} m^{5/4}}{\sqrt{\varepsilon}} \right)$.

References

Ashlagi, I., Monderer, D., and Tennenholtz, M. 2009. Mediators in position auctions. *Games and Economic Behavior 67,* 1, 2–21.

Beckmann, M. J., McGuire, C., and Winsten, C. 1956. *Studies in the economics of transportation.* Yale University Press.

Bhaskar, U., Ligett, K., Schulman, L. J., and Swamy, C. 2014. Achieving target equilibria in network routing games without knowing the latency functions. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on.* IEEE, 31–40.

Caragiannis, I., Kaklamanis, C., and Kanellopoulos, P. 2006. Taxes for linear atomic congestion games. In *Algorithms–ESA 2006.* Springer, 184–195.

Cole, R., Dodis, Y., and Roughgarden, T. 2003. How much can taxes help selfish routing? In *Proceedings of the 4th ACM conference on Electronic commerce.* ACM, 98–107.

Cummings, R., Kearns, M., Roth, A., and Wu, Z. S. 2014. Privacy and truthful equilibrium selection for aggregative games. *CoRR abs/1407.7740.*

Dwork, C., McSherry, F., Nissim, K., and Smith, A. 2006. Calibrating noise to sensitivity in private data analysis. In *TCC '06.* 265–284.

DWORK, C., ROTHBLUM, G. N., AND VADHAN, S. 2010. Boosting and differential privacy. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. 5160.

FLEISCHER, L. 2005. Linear tolls suffice: New bounds and algorithms for tolls in single source networks. *Theoretical Computer Science 348,* 2, 217–225.

FLEISCHER, L., JAIN, K., AND MAHDIAN, M. 2004. Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*. IEEE, 277–285.

FOTAKIS, D., KARAKOSTAS, G., AND KOLLIOPOULOS, S. G. 2010. On the existence of optimal taxes for network congestion games with heterogeneous users. In *Algorithmic Game Theory*. Springer, 162–173.

FREUND, Y. AND SCHAPIRE, R. 1996. Game theory, on-line prediction and boosting. 325–332.

HSU, J., HUANG, Z., ROTH, A., ROUGHGARDEN, T., AND WU, Z. S. 2014a. Private matchings and allocations. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*. 21–30.

HSU, J., HUANG, Z., ROTH, A., AND WU, Z. S. 2014b. Jointly private convex programming. *arXiv preprint arXiv:1411.0998*.

KANNAN, S., MORGENSTERN, J., ROTH, A., AND WU, Z. S. 2015. Approximately stable, school optimal, and student-truthful many-to-one matchings (via differential privacy). In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*. 1890–1903.

KARAKOSTAS, G. AND KOLLIOPOULOS, S. G. 2004. Edge pricing of multicommodity networks for heterogeneous selfish users. In *FOCS*. Vol. 4. 268–276.

KEARNS, M., PAI, M., ROTH, A., AND ULLMAN, J. 2014. Mechanism design in large games: Incentives and privacy. In *Proceedings of the 5th ACM SIGact Innovations in Theoretical Computer Science (ITCS)*.

MCSHERRY, F. AND TALWAR, K. 2007. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*. 94–103.

MONDERER, D. AND SHAPLEY, L. 1996. Potential games. *Games and Economic Behavior 14,* 1, 124–143.

MONDERER, D. AND TENNENHOLTZ, M. 2003. k-implementation. In *Proceedings of the 4th ACM conference on Electronic commerce*. ACM, 19–28.

MONDERER, D. AND TENNENHOLTZ, M. 2009. Strong mediated equilibrium. *Artificial Intelligence 173,* 1, 180–195.

NISSIM, K., SMORODINSKY, R., AND TENNENHOLTZ, M. 2012. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM, 203–213.

PELEG, B. AND PROCACCIA, A. D. 2010. Implementation by mediated equilibrium. *International Journal of Game Theory 39,* 1-2, 191–207.

Raghavan, P. and Thompson, C. D. 1987. Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica 7,* 4, 365–374.

Rogers, R. M. and Roth, A. 2014. Asymptotically truthful equilibrium selection in large congestion games. In *ACM Conference on Economics and Computation, EC '14, Stanford , CA, USA, June 8-12, 2014*. 771–782.

Rozenfeld, O. and Tennenholtz, M. 2007. Routing mediators. In *IJCAI*. 1488–1493.

Slater, M. 1959. Lagrange multipliers revisited. Cowles Foundation Discussion Papers 80, Cowles Foundation for Research in Economics, Yale University.

Swamy, C. 2007. The effectiveness of stackelberg strategies and tolls for network congestion games. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1133–1142.

Zinkevich, M. 2003. Online convex programming and generalized infinitesimal gradient ascent. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA*. 928–936.

A Tools for Differential Privacy

In this section we review the necessary privacy definitions and tools needed for our results. Throughout, let $\mathbf{s} = (s_1, \ldots, s_n) \in \mathcal{S}^n$ be a *database* consisting of n elements from a domain $\mathcal{S}$. In keeping with our game theoretic applications, we refer to the elements $s_1, \ldots, s_n$ as *types* and each type belongs to a *player* $i \in [n]$.

We first state a general lemmas about differential privacy.

Lemma A.1 (Post-Processing Dwork et al. (2006))**.** *Given a mechanism $\mathcal{M} : \mathcal{S}^n \to \mathcal{O}$ and some (possibly randomized) function $p : \mathcal{O} \to \mathcal{O}'$ that is independent of the players' types $\mathbf{s} \in \mathcal{S}^n$, if $\mathcal{M}(\mathbf{s})$ is (ε, δ)-differentially private then $p(\mathcal{M}(\mathbf{s}))$ is (ε, δ)-differentially private.*

A.1 The Laplace Mechanism

We will use the Laplace Mechanism, which was introduced by Dwork et al Dwork et al. (2006) to answers a vector-valued query $f : \mathcal{S}^n \to \mathbb{R}^k$.

The Laplace Mechanism depends on the notation of *sensitivity*—how much a function can change when a single entry in its input is altered.

Definition A.2 (Sensitivity)**.** The sensitivity Δf of a function $f : \mathcal{S}^n \to \mathbb{R}^k$ is defined as

$$\Delta f = \max_{i \in [n], (s_i, \mathbf{s}_{-i}) \in \mathcal{S}^n, s_i' \in \mathcal{S}} \left\{ ||f(s_i, \mathbf{s}_{-i}) - f(s_i', \mathbf{s}_{-i})||_1 \right\}.$$

Algorithm 7 Laplace Mechanism Dwork et al. (2006)

```
Input: : Database $\mathbf{s} \in \mathcal{S}^n$, query $f : \mathcal{S}^n \to \mathbb{R}^k$, and privacy parameter $\varepsilon$.
  procedure $\mathcal{M}_L(\mathbf{s}, f, \varepsilon)$
      Set $\hat{a} = f(\mathbf{s}) + Z$      $Z = (Z_1, \cdots, Z_k)$ and $Z_i \sim \mathrm{Lap}(\Delta f/\varepsilon)$
  return $\hat{a}$.
  end procedure
```

Lemma A.3 (Dwork et al. (2006))**.** *The Laplace Mechanism $\mathcal{M}_L$ is ε-differentially private.*

Lemma A.4 (Dwork et al. (2006))**.** *The Laplace Mechanism $\mathcal{M}_L(\mathbf{s}, f, \varepsilon)$ produces output $\hat{a}$ such that with probability at least $1 - \beta$ we have*

$$||f(\mathbf{s}) - \hat{a}||_\infty \leq \log\left(\frac{k}{\beta}\right)\left(\frac{\Delta f}{\varepsilon}\right)$$

A.2 The Exponential Mechanism

We now present an algorithm introduced by McSherry and Talwar (2007) that is differentially private called the exponential mechanism. Let us assume that we have some finite outcome space $\mathcal{O}$ and a quality score $q : \mathcal{S}^n \times \mathcal{O} \to \mathbb{R}$ that tells us how good the outcome $o \in \mathcal{O}$ is with the given database $\mathbf{s} \in \mathcal{S}^n$. We define the sensitivity of q as the maximum over $o \in \mathcal{O}$ of the sensitivity of $q(\cdot; o)$. Specifically,

$$\Delta q = \max_{o \in \mathcal{O}, \mathbf{s}, \mathbf{s}' \in \mathcal{S}^n} \left\{ |q(\mathbf{s}, o) - q(\mathbf{s}', o)| \right\} \qquad \text{for neighboring } \mathbf{s}, \mathbf{s}'$$

Algorithm 8 Exponential Mechanism McSherry and Talwar (2007)

```
Input: : Database $\mathbf{s} \in \mathcal{S}^n$, quality function $q : \mathcal{S}^n \times \mathcal{O} \to \mathbb{R}$, and privacy parameter $\varepsilon$.
  procedure $\mathcal{M}_E(\mathbf{s}, q, \varepsilon)$
      Output $o \in \mathcal{O}$ with probability proportional to
          $\exp\left(\frac{\varepsilon q(\mathbf{s}, o)}{2\Delta q}\right)$
  end procedure
```

Lemma A.5 (McSherry and Talwar (2007))**.** *The Exponential Mechanism $\mathcal{M}_E$ is ε-differentially private.*

We then define the highest possible quality score with database d to be $OPT_q(\mathbf{s}) = \max_{o \in O}\{q(\mathbf{s}, o)\}$. We then obtain the following proposition that tells us how close we are to the optimal quality score.

Lemma A.6 (McSherry and Talwar (2007))**.** *We have the following utility guarantee from the Exponential Mechanism $\mathcal{M}_E$: with probability at least $1 - \beta$ and every $t > 0$,*

$$q(\mathbf{s}, \mathcal{M}_E(\mathbf{s}, q, \varepsilon)) \geq OPT_q(\mathbf{s}) - \frac{2\Delta q}{\varepsilon}(\log |O| + t)$$

A.3 Composition Theorems

Now that we have given a few differentially private algorithms, we want to show that differentially private algorithms can compose "nicely" to get other differentially private algorithms. We will need to use two composition theorems later in this paper. The first shows that the privacy parameters add when we compose two differentially private mechanisms, and the second from Dwork et al. (2010) gives a better composition guarantee when using many adaptively chosen mechanisms.

Lemma A.7. *If we have one mechanism $M_1 : \mathcal{S}^n \to O$ that is $(\varepsilon_1, \delta_1)$-differentially private, and another mechanism $M_2 : \mathcal{S}^n \times O \to R$ is $(\varepsilon_2, \delta_2)$-differentially private in its first component, then $M : \mathcal{S}^n \to R$ is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$ differentially private where*

$$M(\mathbf{s}) = M_2(\mathbf{s}, M_1(\mathbf{s})).$$

If we were to only consider the previous composition theorem, then the composition of m mechanisms that are ε-differentially private mechanisms would lead to a $m\varepsilon$-differentially private mechanism. However, the next theorem says that we can actually get (ε', δ)-differential privacy where $\varepsilon' = O(\sqrt{m}\varepsilon)$ if we allow for a small $\delta > 0$. This theorem also holds under the threat of an adversary that uses an adaptively chosen sequence of differentially private mechanisms so that each can use the outputs of the past mechanisms and different datasets that may or may not include an individual's data. See Dwork et al. (2010) for further details.

Lemma A.8 (m-Fold Adaptive Composition Dwork et al. (2010))**.** *Fix $\delta > 0$. The class of (ε', δ') differentially private mechanisms satisfies $(\varepsilon, m\delta' + \delta)$ differential privacy under m-fold adaptive composition for*

$$\varepsilon' = \frac{\varepsilon}{\sqrt{8m \log(1/\delta)}}.$$

We also include a proof of Lemma 3.3, the composition of a differentially private algorithm with another joint differentially private algorithm is differentially private.

Proof of Lemma 3.3. Let $S \subseteq O$, $i \in [n]$, and consider data $\mathbf{s} \in \mathcal{S}^n$ and $\mathbf{s}' = (s_i', \mathbf{s}_{-i})$ for $s_i' \in \mathcal{S}$. We have

$$\begin{aligned} \mathbb{P}\left[M(\mathbf{s}) \in S\right] &= \int_{\mathcal{X}^n} \mathbb{P}\left[M_D(\mathbf{x}) \in S\right] \cdot \mathbb{P}\left[M_J(\mathbf{s}) = \mathbf{x}\right] d\mathbf{x} \\ &= \int_{\mathcal{X}^{n-1}} \left[\int_{\mathcal{X}} \mathbb{P}\left[M_D(x_i, \mathbf{x}_{-i}) \in S\right] \cdot \mathbb{P}\left[M_J(\mathbf{s}) = \mathbf{x}\right] dx_i\right] d\mathbf{x}_{-i} \end{aligned}$$

We then use the fact that, since M_D satisfies ε_D-differential privacy, we know that for any fixed $x_i' \in \mathcal{X}$, it holds that $\mathbb{P}\left[M_D(x_i, \mathbf{x}_{-i}) \in S\right] \leq \min\{e^{\varepsilon_D} \cdot \mathbb{P}\left[M_D(x_i', \mathbf{x}_{-i}) \in S\right], 1\}$. We let $P_{x_i', \mathbf{x}_{-i}}$ denote the RHS of this inequality.

$$\begin{aligned} \mathbb{P}\left[M(\mathbf{s}) \in S\right] &\leq \int_{\mathcal{X}^{n-1}} \left[\int_{\mathcal{X}} P_{x_i', \mathbf{x}_{-i}} \cdot \mathbb{P}\left[M_J(\mathbf{s}) = \mathbf{x}\right] dx_i\right] d\mathbf{x}_{-i} \\ &= \int_{\mathcal{X}^{n-1}} P_{x_i', \mathbf{x}_{-i}} \cdot \mathbb{P}\left[M_J(\mathbf{s})_{-i} = \mathbf{x}_{-i}\right] d\mathbf{x}_{-i} \end{aligned}$$

Now we use the fact that, since M_J satisfies (ε_J, δ)-joint differential privacy, we have the inequality $\mathbb{P}\left[M_J(\mathbf{s})_{-i} = \mathbf{x}_{-i}\right] \leq e^{\varepsilon_J} \cdot \mathbb{P}\left[M_J(\mathbf{s}')_{-i} = \mathbf{x}_{-i}\right] + \delta = (\int_{\mathcal{X}} e^{\varepsilon_J} \cdot \mathbb{P}\left[M_J(\mathbf{s}') = \mathbf{x}\right] dx_i) + \delta$.

$$\begin{aligned} \mathbb{P}\left[M(\mathbf{s}) \in S\right] &\leq \int_{\mathcal{X}^{n-1}} P_{x_i', \mathbf{x}_{-i}} \cdot \left[\int_{\mathcal{X}} e^{\varepsilon_J} \cdot \mathbb{P}\left[M_J(\mathbf{s}') = \mathbf{x}\right] dx_i + \delta\right] d\mathbf{x}_{-i} \\ &\leq e^{\varepsilon_J} \cdot \int_{\mathcal{X}^{n-1}} P_{x_i', \mathbf{x}_{-i}} \cdot \left[\int_{\mathcal{X}} \mathbb{P}\left[M_J(\mathbf{s}') = \mathbf{x}\right] dx_i\right] d\mathbf{x}_{-i} + \delta \cdot \int_{\mathcal{X}^{n-1}} P_{x_i', \mathbf{x}_{-i}} d\mathbf{x}_{-i} \end{aligned}$$

Using our definition of $P_{x_i', \mathbf{x}_{-i}}$ (using the first term in the min for the first term above and the second term of the min for the second term above) we can simplify as follows.

$$\mathbb{P}\left[M(\mathbf{s}) \in S\right] \leq e^{\varepsilon_D + \varepsilon_J} \cdot \int_{\mathcal{X}^n} \mathbb{P}\left[M_D(x_i', \mathbf{x}_{-i}) \in S\right] \mathbb{P}\left[M_J(\mathbf{s}') = \mathbf{x}\right] d\mathbf{x} + \delta$$

Again we can apply the fact that, since M_D is ε_D-differentially private, for every $x_i \in \mathcal{X}$, we have that $\mathbb{P}\left[M_D(x_i', \mathbf{x}_{-i}) \in S\right] \leq e^{\varepsilon_D} \cdot \mathbb{P}\left[M_D(\mathbf{x}) \in S\right]$.

$$\begin{aligned} \mathbb{P}\left[M(\mathbf{s}) \in S\right] &\leq e^{2\varepsilon_D + \varepsilon_J} \cdot \int_{\mathcal{X}^n} \mathbb{P}\left[M_D(\mathbf{x}) \in S\right] \mathbb{P}\left[M_J(\mathbf{s}') = \mathbf{x}\right] d\mathbf{x} + \delta \\ &= e^{2\varepsilon_D + \varepsilon_J} \cdot \mathbb{P}\left[M(\mathbf{s}') \in S\right] + \delta \end{aligned}$$

Since this bound holds for every neighboring pair $\mathbf{s}, \mathbf{s}'$, we have proven the lemma. □

B Bounding the Number of Unsatisfied Players

We seek to bound the number of players that are approximately unsatisfied w.r.t. congestion $\hat{\mathbf{y}}$ in the approximately optimal flow $\mathbf{x}^\bullet$ under the routing game $\Gamma^{\hat{\tau}}$, where $\hat{\mathbf{y}}$ is the perturbed version of congestion $\mathbf{y}^\bullet = \sum_i \mathbf{x}_i^\bullet$. First, we give a way to bound the number of unsatisfied players for any approximately optimal flow in the routing game $\Gamma^{\tau^*} = (G, \ell + \tau^*, \mathbf{s})$ that uses the functional marginal-cost tolls $\tau^*(\cdot)$ given in (6).

Lemma B.1. *Let $\rho > 0$ and $\mathbf{x}^\bullet$ be an α-approximately optimal flow in the routing game Γ. Then the number of $\zeta_1(\rho)$-unsatisfied players in Γ^{τ^*} with respect to congestion $\mathbf{y}^\bullet = \sum_{i=1}^n \mathbf{x}_i^\bullet$ is bounded by $n\alpha/\rho$ where*

$$\zeta_1(\rho) = \rho + 4mn\gamma\alpha/\rho$$

Proof. Let $\mathbf{x}$ be any flow in $\mathcal{F}(\mathbf{s})$. Consider the following ρ-best response dynamics: while there exists some ρ-unsatisfied agent i (w.r.t. the true congestion $\sum_i \mathbf{x}_i$), let i make a deviation that decreases her cost the most. Recall that we write $\mathrm{OPT}(\mathbf{s})$ as the optimal value for the routing game Γ. Note that in the tolled routing game Γ^{τ^*}, the potential function Ψ given in (7) satisfies $\Psi(\mathbf{x}) = n \cdot \phi(\mathbf{x})$.

Note that $\mathbf{x}^\bullet$ is an α-approximately optimal flow, so

$$\mathrm{OPT}(\mathbf{s}) \leq \frac{1}{n} \cdot \Psi(\mathbf{x}^\bullet) \leq \mathrm{OPT}(\mathbf{s}) + \alpha.$$

Since each deviation a player made in the dynamics decreases the potential function $\Psi(\mathbf{x})$ by at least ρ, ρ-best response dynamics in game Γ^{τ^*} starting with flow $\mathbf{x}^\bullet$ will terminate after at most $n\alpha/\rho$ iterations. The resulting flow $\hat{\mathbf{x}}$ has all agents ρ-satisfied. In the process, the congestion of each edge might have increased or decreased by at most $n\alpha/\rho$. For each edge $e \in E$, the change in latency is bounded using our γ-Lipschitz condition

$$|\ell_e(y_e) - \ell_e(y_e')| \leq n\gamma\alpha/\rho.$$

Furthermore, the edge toll is also γ-Lipschitz

$$\begin{aligned} |\tau_e^*(y_e) - \tau_e^*(y_e')| &= |(y_e - 1)(\ell_e(y_e) - \ell(y_e - 1)) - (y_e' - 1)(\ell_e(y_e') - \ell(y_e' - 1))| \\ &\leq \gamma |(y_e - 1) - (y_e' - 1)| \leq n\gamma\alpha/\rho. \end{aligned}$$

For the agents that did not deviate in the dynamics, their cost is changed by at most $2mn\gamma\alpha/\rho$. Since they are ρ-satisfied at the end of the dynamics, this means they were $(\rho + 4m\gamma n\alpha/\rho)$-satisfied in the beginning of the process.[4] Since the ρ-best response dynamics lasts for $n\alpha/\rho$ rounds, there are at most $n\alpha/\rho$ number of agents that deviate in the dynamics. □

Based on Lemma B.1, we can now bound the number of approximately unsatisfied players when we impose constant tolls $\tau' = \tau^*(\mathbf{y}^\bullet)$ instead of functional tolls on the edges.

Lemma B.2. *Let $\rho > 0$, $\mathbf{x}^\bullet$ be an α-approximately optimal flow in the routing game Γ, and $\tau' = \tau^*(\sum_i \mathbf{x}_i^\bullet)$ be the vector of constant tolls. Then, the number of $\zeta_2(\rho)$-unsatisfied players with respect to $\mathbf{y}^\bullet = \sum_i \mathbf{x}_i^\bullet$ in the routing game $\Gamma^{\tau'}$ is bounded by $n\alpha/\rho$, where*

$$\zeta_2(\rho) = \rho + 4m\gamma n\alpha/\rho + 2m\gamma.$$

Proof. Let player i be a $\zeta_1(\rho)$-satisfied player in flow $\mathbf{x}^\bullet$ under the routing game Γ^{τ^*}. Now we argue that he should also be $\zeta_2(\rho)$-satisfied under the game $\Gamma^{\tau'}$. Suppose not. Then there exists a route $\mathbf{x}_i'$ for player i that can decrease the cost by more than $\zeta_2(\rho)$ under $\Gamma^{\tau'}$. Now consider the same deviation in game Γ^{τ^*}. Since the functional toll on each edge can change by at most γ, we know that player i's costs in Γ^{τ^*} and $\Gamma^{\tau'}$ differ by at most $2m\gamma$. This implies that the deviation $\mathbf{x}_i'$ in game Γ^{τ^*} could gain him more than $\zeta_1(\rho)$ since $\zeta_2(\rho) - \zeta_1(\rho) = 2m\gamma$.

[4] While the same path agent i is taking might have cost lowered by $2m\gamma n\alpha/\rho$ in the dynamics, any alternate (s_i, t_i)-path might have increased its cost by $2m\gamma n\alpha/\rho$.

From Lemma B.1, we know that the number of $\zeta_1(\rho)$-unsatisfied players under the routing game Γ^{τ^*} is bounded by $n\alpha/\rho$. Therefore, we know that the number of $\zeta_2(\rho)$-unsatisfied players under $\Gamma^{\tau'}$ is also bounded by $n\alpha/\rho$. □

Combining the previous two lemmas, we could now bound the number of unsatisfied players in $\Gamma^{\hat{\tau}}$ with respect to $\mathbf{y} = \sum_{i=1}^n \mathbf{x}_i$ with the differentially private constant tolls $\hat{\tau}$.

Lemma B.3. *Let $\rho, \varepsilon > 0$ and $\mathbf{x}^\bullet$ be an α-approximately optimal flow in the routing game Γ. Let $\hat{\tau} = \tau^*(\hat{\mathbf{y}})$ where $\hat{\mathbf{y}} = \texttt{P-CON}(\mathbf{x}^\bullet, \varepsilon)$. Then with probability at least $1 - \beta$, the number of $\zeta_\varepsilon(\rho)$-unsatisfied players in $\Gamma^{\hat{\tau}}$ with respect to $\mathbf{y}^\bullet = \sum_{i=1}^n \mathbf{x}_i^\bullet$ is bounded by α/ρ, where*

$$\zeta_\varepsilon(\rho) = \zeta_2(\rho) + 4\gamma m^2 \log(m/\beta)/\varepsilon. \tag{21}$$

Proof. From standard bounds on the tails of the Laplace distribution (Lemma A.4), we have the following except with probability β:

$$\max_e \left| \hat{y}_e - \sum_i x_{i,e}^\bullet \right| \leq \frac{2m}{\varepsilon} \cdot \log\left(\frac{m}{\beta}\right)$$

We now condition on this level of accuracy. Since the toll function $\tau^*(\cdot)$ is γ-Lipschitz and $\tau' = \tau^*(\mathbf{y}^\bullet)$, we have

$$\max_e \left| \hat{\tau}_e - \tau'_e \right| \leq \frac{2m\gamma}{\varepsilon} \cdot \log\left(\frac{m}{\beta}\right) \equiv \nu_\varepsilon$$

Therefore a player's cost for taking the same route may increase by as much as $m\nu_\varepsilon$. Further, the cost for an alternative route may decrease by at most the same amount. Thus, each of $\zeta_2(\rho)$-satisfied players under the flow $\mathbf{x}^\bullet$ in Γ^{τ^*} remain $(\zeta_2(\rho) + 2m\nu_\varepsilon)$-satisfied in game $\Gamma^{\hat{\tau}}$. By Lemma B.2, we know that the number of $\zeta_2(\rho)$-unsatisfied players in $\Gamma^{\tau'}$ is bounded by $n\alpha/\rho$, so the number of $(\zeta_2(\rho) + 2m\nu_\varepsilon)$-unsatisfied players in $\Gamma^{\hat{\tau}}$ is bounded by $n\alpha/\rho$ as well. □

We now consider what happens when instead of allowing players to best respond given the exact congestion $\mathbf{y} = \sum_{i=1}^n \mathbf{x}_i$, we instead let them best respond given a private and perturbed version of the congestion. The following general lemma will be useful, which relates to unsatisfied players in two different congestions that are close.

Lemma B.4. *Let Γ be a routing game, and $\mathbf{x}$ be a flow in Γ. Let $\mathbf{y}$ and $\mathbf{y}'$ such that $\|\mathbf{y} - \mathbf{y}'\|_\infty \leq b$. Then for any number $\zeta > 0$, the set of ζ-satisfied players in $\mathbf{x}$ with respect to $\mathbf{y}$ are also ζ'-satisfied with respect to $\mathbf{y}'$, where*

$$\zeta' = \zeta + 2m\gamma b.$$

Proof. The proof follows from the same analysis in the proof of Lemma B.3 □

From the analysis of Lemma B.3, we know that $\|\hat{\mathbf{y}} - \mathbf{y}^\bullet\|_\infty \leq \frac{2m}{\varepsilon} \cdot \log\left(\frac{m}{\beta}\right)$, so by instantiating Lemma B.3 with $\rho = 2\sqrt{m\gamma n\alpha}$ and combining with the result of Lemma B.4, we recover the bound in Lemma 3.6, that is the number of $\hat{\zeta}_\varepsilon$-unsatisfied players w.r.t. congestion $\hat{\mathbf{y}}$ in $\mathbf{x}^\bullet$ and game $\Gamma^{\hat{\tau}}$ is bounded by $\sqrt{n\alpha/4m\gamma}$ with

$$\hat{\zeta}_\varepsilon = 4\sqrt{m\gamma n\alpha} + 8\gamma m^2 \log(m/\beta)/\varepsilon.$$